

Université de Montréal

**Étude structurale du mode de liaison des protéines
Whirly de plantes à l'ADN monocaténaire**

par

Laurent Cappadocia

Département de biochimie

Faculté de médecine

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de Philosophiæ Doctor (Ph. D.)
en biochimie

Décembre 2010

© Laurent Cappadocia, 2010

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée :

Étude structurale du mode de liaison des protéines Whirly de plantes à l'ADN
monocaténaire

présentée par :
Laurent Cappadocia

a été évaluée par un jury composé des personnes suivantes :

Pascale Legault, président-rapporteur
Normand Brisson, directeur de recherche
Jurgen Sygusch, co-directeur
Jacques Archambault, membre du jury
François Belzile, examinateur externe
M. Ram Sairam, représentant du doyen de la FES

Résumé

Les plantes doivent assurer la protection de trois génomes localisés dans le noyau, les chloroplastes et les mitochondries. Si les mécanismes assurant la réparation de l'ADN nucléaire sont relativement bien compris, il n'en va pas de même pour celui des chloroplastes et des mitochondries. Or il est important de bien comprendre ces mécanismes puisque des dommages à l'ADN non ou mal réparés peuvent entraîner des réarrangements dans les génomes. Chez les plantes, de tels réarrangements dans l'ADN mitochondrial ou dans l'ADN chloroplastique peuvent conduire à une perte de vigueur ou à un ralentissement de la croissance. Récemment, notre laboratoire a identifié une famille de protéines, les Whirly, dont les membres se localisent au niveau des mitochondries et des chloroplastes. Ces protéines forment des tétramères qui lient l'ADN monocaténaire et qui accomplissent de nombreuses fonctions associées au métabolisme de l'ADN. Chez *Arabidopsis*, deux de ces protéines ont été associées au maintien de la stabilité du génome du chloroplaste. On ignore cependant si ces protéines sont impliquées dans la réparation de l'ADN.

Notre étude chez *Arabidopsis* démontre que des cassures bicaténares de l'ADN sont prises en charge dans les mitochondries et les chloroplastes par une voie de réparation dépendant de très courtes séquences répétées (de cinq à cinquante paires de bases) d'ADN. Nous avons également montré que les protéines Whirly modulent cette voie de réparation. Plus précisément, leur rôle serait de promouvoir une réparation fidèle de l'ADN en empêchant la formation de réarrangements dans les génomes de ces organites. Pour comprendre comment les protéines Whirly sont impliquées dans ce processus, nous avons élucidé la structure cristalline d'un complexe Whirly-ADN. Nous avons ainsi pu montrer que les Whirly lient et protègent l'ADN monocaténaire sans spécificité de séquence. La liaison de l'ADN s'effectue entre les feuillets β de sous-unités contiguës du tétramère. Cette configuration maintient l'ADN sous une forme monocaténaire et empêche son appariement avec des acides nucléiques de séquence complémentaire. Ainsi, les protéines Whirly peuvent empêcher la formation de réarrangements et favoriser une réparation fidèle de l'ADN. Nous avons également montré que, lors de la liaison de très longues séquences d'ADN, les protéines Whirly peuvent s'agencer en superstructures d'hexamères de tétramères, formant

ainsi des particules sphériques de douze nanomètres de diamètre. En particulier, nous avons pu démontrer l'importance d'un résidu lysine conservé chez les Whirly de plantes dans le maintien de la stabilité de ces superstructures, dans la liaison coopérative de l'ADN, ainsi que dans la réparation de l'ADN chez *Arabidopsis*. Globalement, notre étude amène de nouvelles connaissances quant aux mécanismes de réparation de l'ADN dans les organites de plantes ainsi que le rôle des protéines Whirly dans ce processus.

Mots-clés : Organites de plantes, génomes des chloroplastes et des mitochondries, réparation de l'ADN, réarrangements de l'ADN dépendant de très courtes séquences répétées, protéines liant l'ADN monocaténaire, protéines Whirly, oligomérisation des protéines, radiocristallographie.

Abstract

Plants must protect the integrity of three genomes located respectively in the nucleus, the chloroplasts and the mitochondria. Although DNA repair mechanisms in the nucleus are the subject of multiple studies, little attention has been paid to DNA repair mechanisms in chloroplasts and mitochondria. This is unfortunate since mutations in the chloroplast or the mitochondrial genome can lead to altered plant growth and development. Our laboratory has identified a new family of proteins, the Whirlies, whose members are located in plant mitochondria and chloroplasts. These proteins form tetramers that bind single-stranded DNA and play various roles associated with DNA metabolism. In *Arabidopsis*, two Whirly proteins maintain chloroplast genome stability. Whether or not these proteins are involved in DNA repair has so far not been investigated.

Our studies in *Arabidopsis* demonstrate that DNA double-strand breaks are repaired in both mitochondria and chloroplasts through a microhomology-mediated repair pathway and indicate that Whirly proteins affect this pathway. In particular, the role of Whirly proteins would be to promote accurate repair of organelle DNA by preventing the repair of DNA double-strand breaks by the microhomology-dependant pathway. To understand how Whirly proteins mediate this function, we solved the crystal structure of Whirly-DNA complexes. These structures show that Whirly proteins bind single-stranded DNA with low sequence specificity. The DNA is maintained in an extended conformation between the β -sheets of adjacent protomers, thus preventing spurious annealing with a complementary strand. In turn, this prevents formation of DNA rearrangements and favors accurate DNA repair. We also show that upon binding long ssDNA sequences, Whirly proteins assemble into higher order structures, or hexamers of tetramers, thus forming spherical particles of twelve nanometers in diameter. We also demonstrate that a lysine residue conserved among plant Whirly proteins is important for the stability of these higher order structures as well as for cooperative binding to DNA and for DNA repair. Overall, our study elucidates some of the mechanisms of DNA repair in plant organelles as well as the roles of Whirly proteins in this process.

Keywords: Plant organelles, chloroplastic and mitochondrial genomes, DNA repair, DNA rearrangements mediated by very short repeated sequences, single-stranded DNA binding proteins, Whirly proteins, protein oligomerization, X-ray crystallography.

Table des matières

Liste des tableaux.....	xii
Liste des figures.....	xiii
Liste des abréviations	xvi
Remerciements	xviii
 CHAPITRE 1 : Revue de Littérature	 1
1.1 Dynamique des génomes des organites des plantes.....	2
1.1.1 Historique.....	2
1.1.2 Origine endosymbiotique des organites de plantes	3
1.1.3 Organisation du génome des organites de plantes.....	4
1.1.3.1 Organisation du génome mitochondrial des plantes.....	4
1.1.3.2 Protéines influençant la dynamique du génome mitochondrial	7
1.1.3.3 Organisation du génome plastidique des plantes	8
1.1.3.4 Protéines impliquées dans la formation des nucléoïdes	9
1.1.4 Endommagement et réparation du génome des organites de plantes.....	11
1.1.4.1 Facteurs de stress génomique.....	12
1.1.4.2 Quelques conséquences des stress sur les génomes	13
1.1.4.3 Mécanismes de réparation de l'ADN et de maintien de la stabilité des génomes	14
1.2 Les protéines liant les acides nucléiques monocaténares.....	29
1.2.1 Structure des domaines de liaison aux acides nucléiques monocaténares	29
1.2.1.1 Le domaine OB-fold.....	32
1.2.1.2 Le domaine RRM.....	33
1.2.1.3 Le domaine KH.....	34
1.2.1.4 Le domaine Sm-fold.....	34
1.2.1.5 Le domaine PC4.....	35
1.2.1.6 Le domaine Whirly.....	36
1.2.2 Les stratégies de reconnaissance des acides nucléiques monocaténares	37

1.2.2.1 Types d'interactions impliquées dans la liaison des acides nucléiques monocaténaires	37
1.2.2.2 La spécificité de reconnaissance des acides nucléiques monocaténaires	40
1.2.2.3 La coopérativité dans la liaison des acides nucléiques monocaténaires	42
1.2.2.4 La protection ou l'exposition des bases azotées.....	44
1.2.2.5 Dynamisme des protéines liant l'ADN monocaténaire	45
1.2.3 Les protéines liant l'ADN monocaténaire dans les organites de plantes	45
1.3 Les protéines Whirly de plantes.....	46
1.3.1 Les protéines Whirly accomplissent de nombreux rôles reliés au métabolisme des acides nucléiques.	46
1.3.1.1 Régulation transcriptionnelle du gène <i>PR-10a</i>	46
1.3.1.2 Régulation transcriptionnelle du gène <i>KPI</i>	47
1.3.1.3 Régulation de la longueur des télomères.....	48
1.3.1.4 Épissage des ARN	48
1.3.1.5 Maintien de la stabilité du génome chloroplastique.....	49
1.3.2 Phylogénie des protéines Whirly	49
1.3.2.1 TIF1 : un homologue lointain des Whirly	50
1.3.3 Structure des protéines Whirly de plantes et mécanisme de liaison à l'ADN.....	51
1.4 Objectifs du projet de recherche et justification des méthodes employées	55
1.4.1 Objectifs du projet de recherche.....	55
1.4.2 Justification des méthodes employées.....	55
1.4.2.1 Induction de cassures bicaténaires de l'ADN dans les organites	55
1.4.2.2 Élaboration de la structure tridimensionnelle des protéines Whirly par radiocristallographie	56
1.4.2.3 Élaboration de l'agencement quaternaire des protéines Whirly par radiocristallographie et microscopie à force atomique	58

CHAPITRE 2 : Purification, cristallisation et analyse préliminaire par diffraction aux rayons X du domaine Whirly de St-WHY2 en complexe avec de l'ADN monocaténaire..... 61

2.1 Mise en contexte	63
2.2 Abstract	63
2.3 Introduction	63
2.4 Materials and methods	65
2.5 Results and discussion	67
2.6 Acknowledgements	68
2.7 References	69

CHAPITRE 3 : Structures cristallines de complexes Whirly-ADN et leur rôle dans la réparation du génome des organites d'*Arabidopsis*..... 75

3.1 Mise en contexte	77
3.2 Abstract	77
3.3 Introduction	78
3.4 Results	81
3.5 Discussion	89
3.6 Methods	95
3.7 Acknowledgments	100
3.8 References	100

CHAPITRE 4 : Une lysine des protéines Whirly de plantes est nécessaire pour la protection contre les dommages à l'ADN ainsi que pour la formation de superstructures..... 163

4.1 Mise en contexte	165
4.2 Abstract	165
4.3 Introduction	166
4.4 Experimental procedures	169
4.5 Results	171
4.6 Discussion	176
4.7 References	180
4.8 Footnotes	183

CHAPITRE 5 : Discussion	201
5.1 Implications pour la radiocristallographie	202
5.2 Implications pour la biologie structurale des protéines Whirly	203
5.2.1 Un nouveau mode de liaison à l'ADN monocaténaire.....	203
5.2.2 Liaison à l'ADN et à l'ARN	203
5.2.3 Spécificité de liaison à l'ADN monocaténaire	204
5.2.4 Compétition entre l'ADN et le motif acide/aromatique C-terminal	206
5.2.5 Agencement en 24-mères	206
5.2.6 Liaison coopérative à l'ADN	210
5.2.7 Résidus importants	211
5.2.8 Ressemblances et différences structurales	211
5.3 Implications pour la biologie des organites de plantes	212
5.3.1 Élucidation de mécanismes de réparation des cassures bicaténares de l'ADN dans les organites d' <i>Arabidopsis</i>	213
5.3.2 Rôle des protéines Whirly dans la réparation des cassures bicaténares de l'ADN dans les organites d' <i>Arabidopsis</i>	214
5.4 Perspectives et travaux futurs.....	216
5.4.1 Corroboration des résultats.....	216
5.4.2 Rôle des protéines Whirly dans la réplication de l'ADN.....	216
5.4.3 Redondance fonctionnelle entre les protéines Whirly et les autres protéines liant l'ADN monocaténaire.....	217
5.4.4 Modulation de la conformation de l'ADN par les protéines Whirly.....	217
5.4.5 Rôle de l'agencement en 24-mères des protéines Whirly dans la réparation de l'ADN	218
5.4.6 Liaison de métabolites par les protéines Whirly	218
5.5 Conclusion.....	219
 Bibliographie	 221

Liste des tableaux

Tableau 1-I.	Nombre de courtes séquences (de 50 à 1000 paires de bases) répétées dans les génomes des organites des angiospermes.	6
Tableau 1-II.	Nombre de longues séquences (plus de 1000 paires de bases) répétées dans les génomes des organites des angiospermes.	6
Tableau 1-III.	Structure des domaines de liaison aux acides nucléiques monocaténares	30
Tableau 1-IV.	Structure tridimensionnelle des protéines à domaine Whirly	38
Table 2-I.	Diffraction data statistics.....	70
Table 3-I.	Representative DNA rearrangements observed in <i>Arabidopsis</i> organellar genomes.....	109
Table 3-II.	Data collection and refinement statistics.....	110
Table 3-III.	Analysis of the binding of St-WHY2 to four different DNA sequences by EMSA.	111
Table 3-SI.	DNA rearrangement products identified in the mitochondrial and plastid genomes.....	142
Table 3-SII.	Related DNA rearrangements detected in the mitochondrial and plastid genomes.....	156
Table 3-SIII.	DNA rearrangement statistics.....	157
Table 3-SIV.	DNA rearrangement products detected in the mitochondrial and plastid genomes and used as probes in DNA gel blots.	158
Table 3-SV.	Oligonucleotides used in this study.....	159
Table 4-I.	Data collection and refinement statistics.....	185
Table 4-SI.	Statistical analysis of the complementation of the etiolation/variegation phenotype	200
Tableau 5-I.	Spécificité de liaison des protéines Whirly de plantes aux acides nucléiques. Pour chaque protéine Whirly étudiée, la spécificité est indiquée. La méthode ayant servie à l'analyse est également indiquée.....	205

Liste des figures

Figure 1-1.	Schéma illustrant l'évolution de la morphologie des nucléoïdes durant la différenciation des proplastides en chloroplastes.....	9
Figure 1-2.	Exemples d'agents endommageant l'ADN, de dommages à l'ADN et de mécanismes de réparation de l'ADN	12
Figure 1-3.	Mécanisme de réparation de l'ADN par excision de bases chez les mammifères	16
Figure 1-4.	Mécanisme de réparation de l'ADN par excision de nucléotides chez les mammifères.....	20
Figure 1-5.	Mécanisme de correction des mésappariements dans l'ADN chez les mammifères.....	22
Figure 1-6.	Mécanismes de réparation des cassures bicaténares de l'ADN par la voie de la recombinaison homologue ou par la voie de la jonction d'extrémités non-homologues chez les mammifères.....	24
Figure 1-7.	Modes de liaison à l'ADN monocaténaire de la protéine SSB d' <i>Escherichia coli</i>	43
Figure 1-8.	Modèle proposé pour la liaison de la protéine WHY1 de pomme de terre à l'ADN monocaténaire	53
Figure 1-9.	Interaction entre le domaine Whirly de WHY1 et le motif acide/aromatique C-terminal.....	54
Figure 2-1.	Crystals and diffraction pattern of St-WHY1-ERE ₃₂	71
Figure 2-2.	Crystals and diffraction pattern of St-WHY2-ERE ₃₂	72
Figure 2-3.	SDS-PAGE analysis of dissolved crystals.....	73
Figure 3-1.	PCR strategy to monitor DNA duplication/circularization events	111
Figure 3-2.	DNA rearrangements accumulate in plastids following treatment with the gyrase inhibitor ciprofloxacin.....	112
Figure 3-3.	DNA rearrangements accumulate in mitochondria following treatment with the gyrase inhibitor ciprofloxacin.....	114
Figure 3-4.	Crystal structures of St-WHY2 in the free form and bound to ERE ₃₂ at 2.2-Å resolution.....	116

Figure 3-5.	Mechanism of ssDNA binding of Whirly proteins.....	118
Figure 3-6.	WHY2 binds ssDNA with limited sequence specificity.....	120
Figure 3-7.	WHY2 binds single-stranded overhangs, destabilizes dsDNA, and protects ssDNA against nuclease degradation.	122
Figure 3-8.	Model for the repair of organellar double-strand breaks in the absence or presence of Whirly proteins.	124
Figure 3-S1.	Effect of ciprofloxacin treatment on <i>Arabidopsis</i> plant survival	126
Figure 3-S2.	DNA rearrangements accumulate primarily in plastids of plants lacking plastid-targeted Whirlies and in the mitochondria of plants lacking mitochondria-targeted Whirlies	127
Figure 3-S3.	DNA gel blot detection of DNA rearrangements in plastids following plant treatment with ciprofloxacin.....	128
Figure 3-S4.	DNA gel blot detection of DNA rearrangements in mitochondria following plant treatment with ciprofloxacin.....	130
Figure 3-S5.	Analysis of mitochondrial recombination in plants treated with ciprofloxacin	132
Figure 3-S6.	Analysis of mitochondrial recombination in plants treated with novobiocin	133
Figure 3-S7.	Effects of bleomycin on plastid and mitochondria DNA rearrangements of plants lacking plastid- or mitochondria-targeted Whirlies	134
Figure 3-S8.	Effects of hydroxyurea on plastid and mitochondria DNA rearrangements of plants lacking plastid- or mitochondria-targeted Whirlies	136
Figure 3-S9.	Analysis of mitochondrial recombination in WT plants treated with bleomycin or hydroxyurea	138
Figure 3-S10.	Single-stranded DNA is stabilized on the edges of the β -sheets of Whirly proteins	139
Figure 3-S11.	Modeling of a St-WHY2-RNA complex.....	140
Figure 3-S12.	Molecular characterization of why2 lines.....	141
Figure 4-1.	Critical role of K67 in the assembly of tetramers into 24-mers	186
Figure 4-2.	Whirly proteins form 24-mers in vitro upon binding long ssDNA molecules	188

Figure 4-3.	WHY2 K67A mutant does not assemble into 24-mers <i>in vitro</i>	190
Figure 4-4.	Structural comparison of the K67A variant with the nonmutated WHY2 protein	192
Figure 4-5.	Cooperative binding of WHY2 upon binding long ssDNA molecules	193
Figure 4-6.	WHY1 but not WHY1 K91A partially complements ciprofloxacin sensitivity in <i>why1 why3</i> background	194
Figure 4-S1.	Surface representation of the hexameric assembly of Whirly tetramers in the free form structure of WHY2	196
Figure 4-S2.	Size exclusion chromatography of WHY2 and WHY2 K67A in the free form or in complex with a dT ₃₂ oligonucleotide	198
Figure 5-1.	Représentation des résidus importants pour l'interaction avec l'ADN monocaténaire, pour l'interaction avec le motif acide/aromatique C-terminal, pour la stabilisation du tétramère ou la stabilisation du 24-mère	207
Figure 5-2.	Représentation en style surface de protéines s'assemblant en 24-mères	209

Liste des abréviations

AFM :	Atomic Force Microscopy
At-WHY1 :	<i>Arabidopsis thaliana</i> Whirly 1
At-WHY2 :	<i>Arabidopsis thaliana</i> Whirly 2
At-WHY3 :	<i>Arabidopsis thaliana</i> Whirly 3
AtKP1 :	<i>Arabidopsis thaliana</i> Kinesin-Like Protein 1
BIR :	Break-Induced Replication
CND41 :	41-kDa Chloroplast Nucleoid DNA-Binding Protein
CRS1 :	Chloroplast RNA Splicing 1
dsDNA :	Double-Stranded DNA
DSB :	DNA Double-Strand Break
EMSA :	Electrophoretic Mobility Shift Assay
ERE :	Elicitor Response Element
KH :	K Homology
KPRE :	AtKP1-Related Element
MFP1 :	MAR-Binding Filament-Like Protein 1
MHMR :	Microhomology-Mediated DNA Rearrangement
MMBIR :	Microhomology-Mediated Break-Induced Replication
NHEJ :	Non Homologous End-Joining
OB-fold :	Oligonucleotide/Oligosaccharide Binding-fold
OSB :	Organellar Single-Stranded DNA Binding
PC4 :	Positive Cofactor 4
PEND :	Plastid-Envelope DNA-Binding Protein
PR-10a :	Pathogenesis Related 10a
RMSD :	Root-Mean Square Deviation
RNP :	Ribonucleoprotein
RPA :	Replication Protein A
RRM :	RNA Recognition Motif
SE :	Silencing Element
SSA :	Single-Stranded Annealing

SSB :	Single-Stranded DNA-Binding Protein
ssDNA :	Single-Stranded DNA
St-WHY1 :	<i>Solanum tuberosum</i> Whirly 1
St-WHY2 :	<i>Solanum tuberosum</i> Whirly 2
TEBP :	Telomere End Binding Protein
TFIIH :	Transcription Factor IIH
XPC :	<i>Xeroderma Pigmentosum</i> Complementation Group C
Zm-WHY1 :	<i>Zea Mays</i> Whirly 1

Remerciements

Je voudrais remercier mes directeurs de recherche pour m'avoir accueilli dans leur laboratoire et pour m'avoir donné une très grande liberté pour mener à bien ce projet de recherche.

Je tiens à remercier les nombreux membres des laboratoires avec qui j'ai eu la chance de travailler. En particulier, je voudrais remercier les trois autres mousquetaires du laboratoire : Jean-Sébastien Parent, Étienne Lepage et Alexandre Maréchal. Cela a été un réel plaisir de vous voir tous les jours au laboratoire. Vous avez su créer une ambiance joyeuse et délirante qui a permis de traverser les temps plus durs le sourire aux lèvres. Merci à Jean-Sébastien Parent, pour les presque dix années que j'ai passé avec toi au baccalauréat et au laboratoire. Tu as été une vraie source d'inspiration. Merci à Étienne Lepage pour ta grande générosité, ton humour et ton « attitude de gagnant ». Merci à Alexandre Maréchal pour ton sourire contagieux et ta détermination. Un grand merci également à Mathieu Coinçon pour tes nombreuses discussions et ton franc parler. Merci à mes stagiaires Thomas Morse et Éléni Sirokakis. Merci à Éric Zampini et Thierry Vincent, « the next generation ». Finalement, je remercie le personnel du département, en particulier Sylvie Beauchemin. Merci au personnel du synchrotron de Brookhaven. Merci aussi aux professeurs Jim Omichinski et Luc Varin pour leur temps et leur aide précieuse ainsi qu'aux membres de mon jury de thèse.

Un grand merci à mes parents qui m'ont toujours soutenu. Merci à Adam Korzekwa. Finalement, je garde mon remerciement absolu pour ma fiancée, Valérie. Merci pour ton amour et ton soutien tout au long de mon parcours.

CHAPITRE 1

Revue de littérature

1.1 Dynamique des génomes des organites des plantes

1.1.1 Historique

C'est au début du XX^e siècle qu'apparaît l'idée que les organites de plantes contiennent de l'information génétique. Nous sommes alors quelques années après la redécouverte des lois de l'hérédité de Gregor Mendel par Hugo de Vries, Erich von Tschermak et Carl Correns. Ce dernier, qui a continué de travailler sur l'hérédité des plantes, découvre des incohérences dans la transmission de certains caractères chez la Belle de nuit (*Mirabilis jalapa*; pour revue, se référer à (1)). Il s'intéresse aux croisements entre des plantes vertes et des plantes variégues, c'est-à-dire des plantes présentant des secteurs végétatifs de différentes couleurs, et remarque que lorsqu'une plante verte est fécondée avec le pollen d'une plante variégue, la progéniture ne présente jamais de variégation tandis que lorsque qu'une plante variégue est fécondée avec le pollen d'une plante verte, de la variégation est observée dans la progéniture, mais pas selon les rapports mendéliens. Ceci constitue le premier cas documenté de transmission uniparentale de l'information génétique et cette observation sera à l'origine de la théorie de transmission cytoplasmique de l'information génétique.

Il faudra cependant attendre 1962 pour qu'il soit formellement montré que les chloroplastes et mitochondries contiennent de l'ADN. Cette année là, une étude par microscopie électronique démontre que les chloroplastes de l'algue verte *Chlamydomonas* contiennent des filaments qui sont sensibles à un traitement par une DNase. L'année suivante, il est montré par la même technique que les mitochondries contiennent elles aussi de l'ADN. Au cours des années suivantes, plusieurs groupes parviendront à purifier de l'ADN à partir de mitochondries et de chloroplastes de nombreuses espèces (Pour revue, se référer à (2-4)). Cet ADN diffère de l'ADN nucléaire de par sa température de dénaturation et sa composition en G/C.

Vers le milieu des années 70, le développement des techniques de biologie moléculaire permet l'apparition des premières cartes physiques du génome des organites. Il est alors bien accepté que les organites possèdent leur propre génome constitué d'ADN bicaténaire

circulaire. En 1981 on élucide la séquence complète du génome des mitochondries humaines (5); en 1986 celles des génomes des chloroplastes de tabac (*Nicotiana tabacum*; (6)) et d'hépatique (*Marchantia polymorpha*; (7)). Depuis lors, les séquences de très nombreux génomes d'organites ont été obtenus, permettant ainsi d'effectuer des analyses comparatives.

Aujourd'hui, les mécanismes de réplication, recombinaison et réparation de l'ADN des organites de plantes sont encore beaucoup moins bien connus que ceux du noyau. Cette différence est possiblement due à l'absence de nombreux outils qui ont été développés pour étudier le génome nucléaire tels les systèmes rapporteurs de la réplication, recombinaison et réparation de l'ADN et les systèmes d'induction de dommages à l'ADN nucléaire par expression hétérologue d'enzymes de restriction (pour revue, se référer à (8)). La recherche concernant la dynamique du génome des organites est cependant stimulée depuis quelques années par l'identification chez l'humain de liens entre l'accumulation de mutations mitochondriales et le cancer et le vieillissement (9,10). Chez les plantes, la mise en évidence de traits bénéfiques pour l'humain, telle la stérilité male cytoplasmique pour la production d'hybrides (11) ou la variégation pour la production de plantes ornementales (12), constitue un motif d'étude des génomes des mitochondries et des chloroplastes. Nous allons à présent voir l'origine des organites de plantes, comment leurs génomes sont organisés et comment ils se protègent contre des modifications inopinées de leur ADN.

1.1.2 Origine endosymbiotique des organites de plantes

Parallèlement à la découverte des génomes des organites, apparaît la théorie de l'origine endosymbiotique des organites qui suggère que les mitochondries et les plastides sont apparus au cours de l'évolution suite à des phénomènes d'endosymbiose. Il est aujourd'hui bien admis que les mitochondries proviendraient de l'encapsulation d'une α -protéobactérie par un ancêtre des cellules eucaryotes, tandis que l'encapsulation subséquente d'un ancêtre des cyanobactéries aurait quant à elle donné naissance aux plastides (pour revue, se référer à (13)). Après l'établissement des plastides et des mitochondries, le génome de ces organites a été réorganisé. Beaucoup de gènes ont été perdus ou transférés vers le noyau (pour revue, se référer à (14)). Ainsi, les organites de plantes ne possèdent aujourd'hui qu'un génome restreint qui ne code que pour une petite fraction de leur protéome. On estime ainsi que

les chloroplastes n'encodent que cinq à dix pour cent des protéines retrouvées dans les cyanobactéries actuelles, tandis que les mitochondries n'encodent qu'un à trois pour cent des protéines retrouvées dans les α -protéobactéries actuelles (15).

Le transfert de gènes des organites vers le noyau a nécessité l'élaboration d'un système de ciblage permettant aux protéines encodées par ces gènes d'être dirigées et importées dans les organites (pour revue, se référer à (13)). Un système de communication entre le noyau et les organites a également dû être créé de manière à coordonner les activités nucléaires avec celles des organites (16). Le système de ciblage nécessite que les protéines dirigées vers les organites possèdent un signal de localisation vers leur organe cible. Ce signal se manifeste sous forme d'un peptide de transit généralement situé dans la portion N-terminale de la protéine (pour revue, se référer à (13)). Celui-ci sera reconnu par des récepteurs situés au niveau de la membrane externe des organites, ce qui permettra le transport de la protéine à l'intérieur de l'organe. Il est intéressant de noter que, durant le cours de l'évolution des plantes, les gènes codant pour la machinerie d'entretien et d'expression du génome des organites sont entièrement passés sous contrôle nucléaire (17,18) et que de nombreux gènes codant pour des protéines impliquées dans le métabolisme de l'ADN ou de l'ARN dans une organe ont été dupliqués et codirigés vers l'autre organe (19).

1.1.3 Organisation du génome des organites de plantes

1.1.3.1 Organisation du génome mitochondrial des plantes

Le génome mitochondrial des plantes compte parmi les plus grands génomes mitochondriaux. Celui de certaines plantes à fleur peut atteindre de 200 000 à 2 400 000 paires de bases (20), avec un coefficient GC % voisin de 43 (21). Nonobstant leur grande taille, les génomes mitochondriaux des plantes ne codent pas pour beaucoup plus de protéines que les génomes mitochondriaux plus petits d'autres espèces. Ainsi, le génome mitochondrial d'*Arabidopsis*, qui comprend 57 gènes pour une taille de 367 000 paires de bases (22), est vingt fois plus grand que celui de l'humain (37 gènes pour une taille de 16 500 paires de bases (5)) mais contient à peine plus de gènes. Cette différence de taille est imputable à la grande quantité d'ADN non codant présente chez les mitochondries des

plantes. Chaque espèce végétale possède en effet de l'ADN mitochondrial non codant qui lui est propre et que l'on ne retrouve même pas chez des espèces proches (22-27). On estime cependant qu'une partie de cet ADN proviendrait d'un transfert à partir des plastides et du noyau (pour revue, se référer à (21)). Des études par microscopie à fluorescence ont révélé que l'ADN des mitochondries de plantes est organisé en complexes nucléoprotéiques appelés nucléoïdes. Ces nucléoïdes sont le siège d'une activité de synthèse et de transcription de l'ADN (28). En revanche, nous ne connaissons pas encore l'identité des protéines impliquées dans l'assemblage de ces nucléoïdes (29).

L'analyse par restriction du génome mitochondrial des plantes a mené à l'hypothèse que celui-ci serait organisé en un « cercle maître », c'est-à-dire un ADN circulaire qui contiendrait tous les gènes mitochondriaux (30). Des études subséquentes par électrophorèse et microscopie électronique ont cependant révélé que le génome mitochondrial des plantes existerait plutôt sous différentes formes : linéaire, circulaire, branchée ((31,32) et pour revue, se référer à (33)). Ces formes seraient créées par recombinaison d'ADN entre des séquences répétées disséminées dans le génome. On distingue à cet égard trois catégories de séquences répétées (34,35) : les très courtes (moins de 50 nucléotides), les courtes (entre 50 et 1000 nucléotides) et les longues (plus de 1000 nucléotides). L'analyse de la séquence des génomes mitochondriaux des plantes démontre que ceux-ci sont très riches en séquences répétées disséminées (Tableaux 1-I et 1-II). Ainsi, chez *Arabidopsis*, on compte deux longues et 60 courtes séquences répétées (34). On observe fréquemment une recombinaison entre les longues séquences répétées. Ce processus est réversible et donne naissance à des ADN subgénomiques qui ne possèdent qu'une partie des gènes mitochondriaux ((36) et pour revue, se référer à (37)). Des changements dans le ratio des ADN subgénomiques sont appelés des variations substoichiométriques, ou *substoichiometric shifting*, et sont notamment observés suite à un changement de conditions de croissance des plantes (38). Un autre type de recombinaison, moins fréquent, survient entre des séquences répétées de courte ou de très courte taille (34,39,40) et donne lieu à des réarrangements généralement irréversibles du génome (41). Une conséquence de ces réarrangements est que l'organisation du génome mitochondrial (l'ordre des gènes) n'est pas constante pour différentes espèces de plantes. Elle varie non seulement d'une espèce à l'autre, mais parfois au sein d'une même espèce (21,42). En contraste avec ces fréquents réarrangements du génome mitochondrial,

Tableau 1-I. Nombre de courtes séquences (de 50 à 1000 paires de bases) répétées dans les génomes des organites des angiospermes.

Espèce	Nombre de courtes séquences répétées d'ADN		Numéro de référence
	Chloroplaste	Mitochondrie	
<i>Arabidopsis thaliana</i>	0	60	NC_000932 NC_001284
<i>Carica papaya</i>	0	32	NC_010323 NC_012116
<i>Nicotiana tabacum</i>	0	48	NC_001879 NC_006581
<i>Vitis vinifera</i>	1	100	NC_007957 NC_012119
<i>Oryza sativa (indica)</i>	1	61	NC_008155 NC_007886
<i>Sorghum bicolor</i>	2	39	NC_008602 NC_008360
<i>Triticum aestivum</i>	2	54	NC_002762 NC_007579
<i>Zea mays</i>	2	49	NC_001666 NC_007982

Ces séquences ont été identifiées à partir du programme REPUTER. Ce tableau est librement inspiré de celui de (34).

Tableau 1-II. Nombre de longues séquences (plus de 1000 paires de bases) répétées dans les génomes des organites des angiospermes.

Espèce	Nombre de longues séquences répétées d'ADN		Numéro de référence
	Chloroplaste	Mitochondrie	
<i>Arabidopsis thaliana</i>	1	2	NC_000932 NC_001284
<i>Carica papaya</i>	1	1	NC_010323 NC_012116
<i>Nicotiana tabacum</i>	1	3	NC_001879 NC_006581
<i>Vitis vinifera</i>	1	0	NC_007957 NC_012119
<i>Oryza sativa (indica)</i>	1	12	NC_008155 NC_007886
<i>Sorghum bicolor</i>	1	4	NC_008602 NC_008360
<i>Triticum aestivum</i>	1	10	NC_002762 NC_007579
<i>Zea mays</i>	1	4	NC_001666 NC_007982

Ces séquences ont été identifiées à partir du programme REPUTER. Ce tableau est librement inspiré de celui de (34).

on observe que le taux de mutation dans la mitochondrie est beaucoup moins élevé que celui du chloroplaste ou du noyau (43,44).

1.1.3.2 Protéines influençant la dynamique du génome mitochondrial

Les protéines responsables de la réplication, de la recombinaison et de la réparation de l'ADN dans les mitochondries de plantes sont encodées dans le noyau (17) et la suppression de certains gènes nucléaires peut mener à l'apparition de réarrangements dans le génome mitochondrial. Plusieurs exemples sont documentés : (i) le gène *Msh1* d'*Arabidopsis* dont le produit est homologue à la protéine de correction des mésappariements MutS de *Escherichia coli* (45), (ii) le gène *OSB1* (*Organellar Single-Stranded DNA-Binding 1*) dont le produit est une protéine liant l'ADN monocaténaire que l'on retrouve principalement chez les plantes (46), (iii) le gène *RecA3* d'*Arabidopsis* (47) ainsi que le gène *RecA1* de la mousse *Physcomitrella patens* (48) dont les produits sont homologues à la recombinaison RecA de *Escherichia coli*. La suppression de ces protéines entraîne l'apparition de réarrangements médiés par des séquences d'ADN de courte taille dans l'ADN mitochondrial. Une étude récente chez *Arabidopsis* montre même qu'en l'absence de *Msh1*, une trentaine de séquences répétées de taille comprise entre 108 et 556 paires de bases sont utilisées dans le cadre de recombinaisons (49). Plusieurs phénotypes sont associés à la suppression de ces gènes et aux réarrangements qu'ils occasionnent. Chez des plantes d'*Arabidopsis* déficientes pour *Msh1* ou pour *OSB1*, on observe une augmentation du pourcentage de variégation soit dans la plante mutée, soit dans sa descendance (46,50). Des plantes de tomate et de tabac dont l'expression du gène *Msh1* a été réduite développent aussi de la variégation ainsi que de la stérilité mâle cytoplasmique (51). Bien qu'aucun phénotype visible ne soit lié à l'absence de *RecA3* d'*Arabidopsis* (47), des plantes mutantes pour *RecA3* et *Msh1* développent quant à elles des phénotypes sévères (52). Cependant, certaines de ces plantes sont plus tolérantes aux hautes températures, ce qui s'expliquerait par une augmentation de la quantité d'ARN codant pour des protéines impliquées dans les réponses aux stress (52).

1.1.3.3 Organisation du génome plastidique des plantes

Le génome plastidique des plantes possède une taille généralement comprise entre 115 000 et 165 000 paires de bases, avec un coefficient GC % voisin de 37 (15). Il comporte de 110 à 130 gènes dont la majorité code pour des protéines impliquées dans la photosynthèse ou l'expression des gènes. Les autres gènes codent pour des ARN de transfert ou des ARN ribosomiaux. Le génome plastidique des plantes possède une organisation particulière constituée de deux séquences uniques (une longue et une courte) séparées par deux séquences répétées inversées. Ces dernières ont une taille variant généralement entre 20 000 et 30 000 paires de bases bien que certaines espèces de plantes n'en possèdent pas (15). Ces séquences répétées inversées sont le site d'une recombinaison intramoléculaire intense qui donne lieu à la formation d'isomères de structures du génome (53). Bien que la majorité des plantes ne possèdent que peu de courtes séquences répétées disséminées dans leur génome plastidique (Tableaux 1-I et 1-II; (34)), on en dénombre jusqu'à deux cent dans le génome plastidique de certaines algues vertes (54).

L'ordre des gènes chloroplastiques est relativement bien conservé, tant d'un tissu à l'autre que d'une espèce à l'autre (55). Le génome plastidique accumule moins de mutations que le génome nucléaire mais plus que le génome mitochondrial (56). Des événements de délétion ou de duplication d'ADN médiés par de très courtes séquences riches en A/T ont également été observées en comparant les génomes chloroplastiques de différentes graminées (57,58).

On a longtemps pensé que le génome plastidique était circulaire ((59) et pour revue, se référer à (60)). Bien que des ADN linéaires aient été visualisés en microscopie électronique, ceux-ci étaient considérés comme des artefacts dus à la brisure d'ADN circulaire lors de la préparation des échantillons. Il est aujourd'hui admis que le génome plastidique peut adopter différentes formes, y compris des formes circulaires, linéaires et branchées (60,61). Ces formes sont créées lors de la réplication ou de la recombinaison du génome chloroplastique.

Dans chaque chloroplaste, le génome est présent en copies multiples et, globalement, chaque cellule possède entre mille et dix mille copies du génome plastidique (15). Comme dans le cas des mitochondries, l'ADN est généralement organisé en complexes

nucléoprotéiques appelés nucléoïdes. Ces nucléoïdes contiennent des protéines de liaison à l'ADN qui en maintiennent la structure ainsi que des enzymes qui répliquent et transcrivent le génome chloroplastique. Les nucléoïdes de feuilles matures ont un diamètre d'environ 0.2 μm et contiennent chacun une dizaine de copies du génome (62). Leur position au sein du chloroplaste varie en fonction du développement (Figure 1-1 et (29)). Chez les chloroplastes de feuilles matures, ils sont localisés dans le stroma et sont associés à la membrane des thylakoïdes.

1.1.3.4 Protéines impliquées dans la formation des nucléoïdes

Plusieurs études se sont intéressées aux protéines qui co-localisent avec les nucléoïdes et en condensent l'ADN. Chez les algues rouges, un homologue de la protéine HU bactérienne a été découvert (63). Cette protéine, nommée HC, est également présente

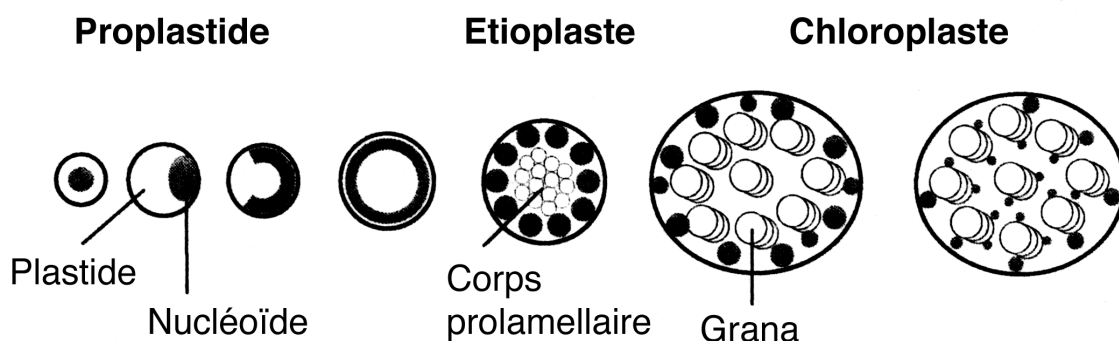


Figure 1-1. Schéma illustrant l'évolution de la morphologie des nucléoïdes durant la différenciation des proplastides en chloroplastes. Dans la graine, les proplastides contiennent un petit nucléoïde localisé au centre du plastide. Durant la germination, le nucléoïde va migrer vers la périphérie du plastide, s'agrandir et prendre la forme d'un anneau. Par la suite, le nucléoïde va se diviser. Durant l'étiolation, les nucléoïdes vont s'aligner autour du corps prolamellaire. Une fois exposé à la lumière, l'étioplaste va se différencier en chloroplaste et certains nucléoïdes vont se diviser et migrer vers le centre du plastide. Cette figure est adaptée de (29) avec la permission de Elsevier.

chez les cyanobactéries, les précurseurs des chloroplastes, mais elle est absente chez les algues vertes ainsi que chez les plantes terrestres (63). Chez ces dernières, quatre protéines impliquées dans le positionnement ou le maintien de la structure des nucléoïdes ont été identifiées : la ferredoxine:sulfite réductase, CND41 (*41-kDa chloroplast nucleoid DNA-binding protein*), MFP1 (*MAR-binding filament-like protein 1*) et PEND (*Plastid-envelope DNA-binding protein*) (64).

La ferredoxine:sulfite réductase (65-67) est une enzyme redox qui participe à la voie d'assimilation du soufre et qui peut se lier à l'ADN mono- ou bicaténaire sans que cette liaison n'affecte ses activités enzymatiques (67). L'ajout de cette protéine à de l'ADN chloroplastique entraîne une condensation de l'ADN, une diminution de la transcription et une diminution de la synthèse d'ADN *in vitro* (68,69). Des plantes de *Nicotiana benthamiana* chez qui le niveau de cette protéine a été diminué présentent des altérations dans la quantité et la structure de leur ADN chloroplastique ainsi que dans la transcription de certains gènes chloroplastiques (70). Ces effets sont probablement dus à l'activité de liaison à l'ADN de la protéine, étant donné que des plantes qui croissent en présence de cystéine, le produit final de la voie d'assimilation du soufre, présentent le même phénotype (70). Bien que le lien exact entre l'activité enzymatique de la ferredoxine:sulfite réductase et son activité de condensation de l'ADN soit inconnu, cette protéine pourrait servir de senseur qui modulerait l'activité de transcription et de réplication des nucléoïdes suite à une fluctuation des niveaux métaboliques (67).

La protéine CND41 a été isolée à partir des nucléoïdes de chloroplastes de tabac (71). Elle est abondamment présente dans les plastides en division et peu présente dans les plastides transcriptionnellement actifs (71). Elle possède un motif hélice-tour-hélice avec une région basique à son extrémité N-terminale qui lui permet de lier l'ADN de manière non-spécifique (71). CND41 possède également une activité enzymatique de protéase à aspartate (72) qui est inhibée lors de la liaison à l'ADN (73). CND1 serait impliqué dans la sénescence. Plus précisément, une diminution de la quantité d'ADN chloroplastique dans les cellules en sénescence entraînerait une relâche de l'inhibition de CND41 et la dégradation concomitante de protéines (73).

MFP1 est une protéine à domaine coiled-coil qui possède un domaine transmembranaire dans sa portion N-terminale et un domaine de liaison à l'ADN dans sa portion C-terminale (74). Initialement découverte dans le noyau (75), cette protéine se localise principalement au niveau des chloroplastes (76) où elle lie l'ADN bicaténaire sans spécificité de séquence et permet l'ancrage des nucléoïdes aux thylakoïdes des chloroplastes matures (74). MFP1 peut être phosphorylée par la protéine CKII, ce qui entraîne une diminution de son affinité pour l'ADN (77).

La protéine PEND est présente chez les plantes terrestres mais absente chez les algues vertes et chez les cyanobactéries, ce qui suggère qu'elle est apparue tardivement au cours de l'évolution (78). Cette protéine, qui forme un homodimère dont chaque sous-unité contient un zipper de leucine basique dans sa portion N-terminale et un domaine transmembranaire dans sa portion C-terminale, lie spécifiquement un élément de huit nucléotides d'ADN (64). Ceci permettrait d'ancrer les nucléoïdes à l'enveloppe interne du plastide au début du développement du chloroplaste (79).

La purification de nucléoïdes ou de complexes similaires tels les chromosomes transcriptionnellement actifs a permis d'identifier d'autres protéines associées aux nucléoïdes (80). On ignore cependant lesquelles de ces protéines sont directement associées à l'ADN ou impliquées dans le maintien de l'organisation tridimensionnelle des nucléoïdes.

1.1.4 Endommagement et réparation du génome des organites de plantes

Des modifications du génome surviennent constitutivement dans la vie des cellules de par leur activité métabolique ou de par leur interaction avec l'environnement (81). Nous allons voir ici quels sont les principaux stress génomiques auxquels les cellules doivent faire face, ainsi que les mécanismes de réparation des génomes (Figure 1-2). Dès que possible nous mettrons en relation ces mécanismes avec ceux connus dans les organites de plantes.

1.1.4.1 Facteurs de stress génomique

Le stress réplcatif de l'ADN

On parle de stress réplcatif lors d'un ralentissement, d'un arrêt ou d'un effondrement des fourches de réplication. Le stress réplcatif peut conduire à l'apparition de lésions dans l'ADN ainsi qu'à de l'instabilité génomique. Plusieurs facteurs peuvent engendrer un stress de réplication : un épuisement des réserves de nucléotides triphosphates nécessaires à la synthèse de l'ADN, un débalancement dans le niveau d'expression des protéines impliquées dans la réplication, un changement dans la fréquence d'initiation de la réplication, des lésions dans l'ADN ou l'inhibition de la réplication par des composés chimiques. Un stress

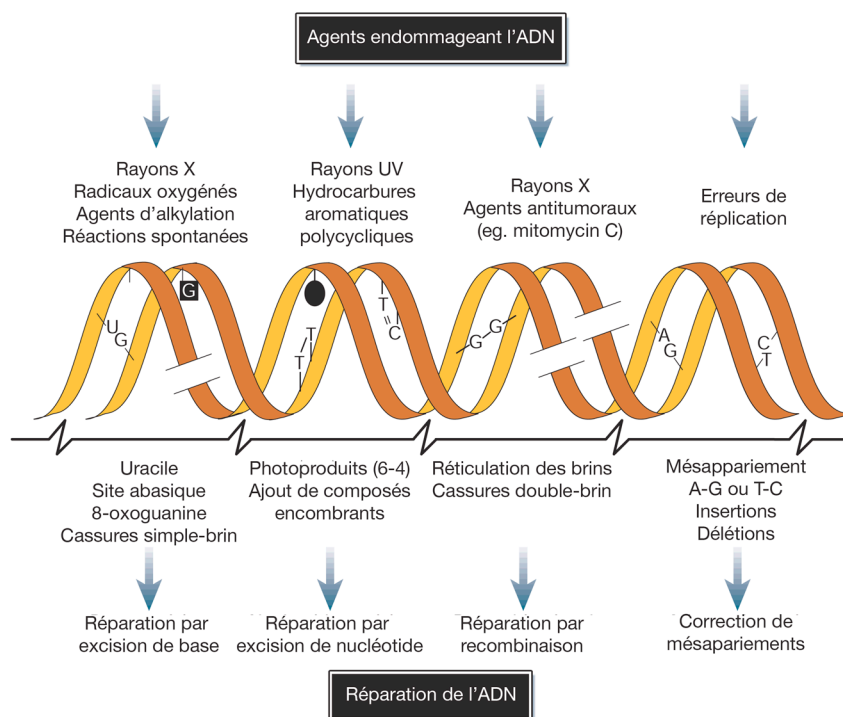


Figure 1-2. Exemples d'agents endommageant l'ADN (en haut), de dommages à l'ADN (au milieu) et de mécanismes de réparation de l'ADN (en bas). Cette figure est adaptée de (81) avec la permission de Macmillan Publishers Ltd (copyright 2001).

réplicatif peut également survenir si la séquence d'ADN à répliquer adopte une structure secondaire stable ou est difficile à dénaturer ou si des protéines y sont fortement liées (82).

Le stress environnemental

Des facteurs environnementaux tels les composés génotoxiques (ex. inhibiteurs de topoisomérases) et le rayonnement électromagnétique de haute énergie (rayons gamma, rayons X ou rayons ultraviolets) peuvent altérer l'ADN (81). Dans le cas du rayonnement électromagnétique, les dommages peuvent être causés de manière directe, par la modification ou le clivage de l'ADN, ou de manière indirecte, par formation de radicaux libres et d'espèces réactives qui peuvent à leur tour endommager l'ADN.

Le stress métabolique

Des sous-produits du métabolisme cellulaire comme les espèces oxygénées réactives (anions superoxydes, radicaux libres et peroxyde d'hydrogène) peuvent entraîner une altération des bases azotées ou introduire des cassures dans l'ADN. Des espèces oxygénées réactives sont régulièrement formées au sein des organites de plantes, étant donné que ces organites sont le siège d'une intense activité métabolique (83,84).

Les modifications spontanées

En condition physiologique, il peut y avoir clivage ou modification spontanée de l'ADN. Les bases azotées sont particulièrement sensibles aux modifications spontanées. De plus, étant donné que certaines bases ne diffèrent que par quelques groupements chimiques, la perte ou la modification de ces groupements peut rapidement amener à un changement de sens de l'information génétique. Ainsi, la désamination d'une cytosine peut transformer cette base en uracile. Similairement, une adénine peut devenir une hypoxanthine, une guanine devenir une xanthine et une 5-méthylcytosine devenir une thymine.

1.1.4.2 Quelques conséquences des stress sur les génomes

Les stress peuvent avoir différentes conséquences pour l'ADN. Il peut y avoir (i) modification chimique de la structure d'une base azotée ou d'un nucléotide, (ii) perte de

groupements fonctionnels de l'ADN comme une base azotée, et *(iii)* cassure de l'ADN. On distingue à cet égard deux types de cassures : les cassures monocaténares et les cassures bicaténares. Ces dernières sont les plus nocives pour l'ADN (pour revue, se référer à (85)). Elles surviennent lorsque les deux brins de l'ADN sont clivés simultanément ou lorsque deux cassures monocaténares apparaissent à proximité l'une de l'autre ou lorsqu'une fourche de réplication rencontre une cassure monocaténaire d'ADN. Chez la levure, une seule cassure bicaténaire non réparée peut conduire à la mort cellulaire. Les cassures bicaténares mal réparées peuvent quant à elles conduire à de l'instabilité génomique. Paradoxalement, des cassures bicaténares sont parfois volontairement générées par les cellules. Celles-ci permettent notamment d'accroître la variabilité génétique dans le cadre de la recombinaison méiotique (86) ou d'augmenter la variabilité d'épitopes dans le cadre du développement du système immunitaire.

Des dommages non réparés ou réparés de manière fautive peuvent provoquer des altérations dans le génome. D'une part, il peut y avoir modification dans le nombre de copies du génome (niveau de ploïdie) ou dans le nombre de chromosomes. D'autre part, il peut y avoir changement dans la structure du génome (insertion, délétion, transposition), dans le nombre de copies d'une séquence d'ADN, ou dans la séquence d'ADN (polymorphismes de nucléotides). Pour se protéger contre ces événements, les organismes ont mis en place des mécanismes leur permettant de reconnaître et réparer adéquatement les dommages causés à l'ADN.

1.1.4.3 Mécanismes de réparation de l'ADN et de maintien de la stabilité des génomes

La tolérance aux lésions

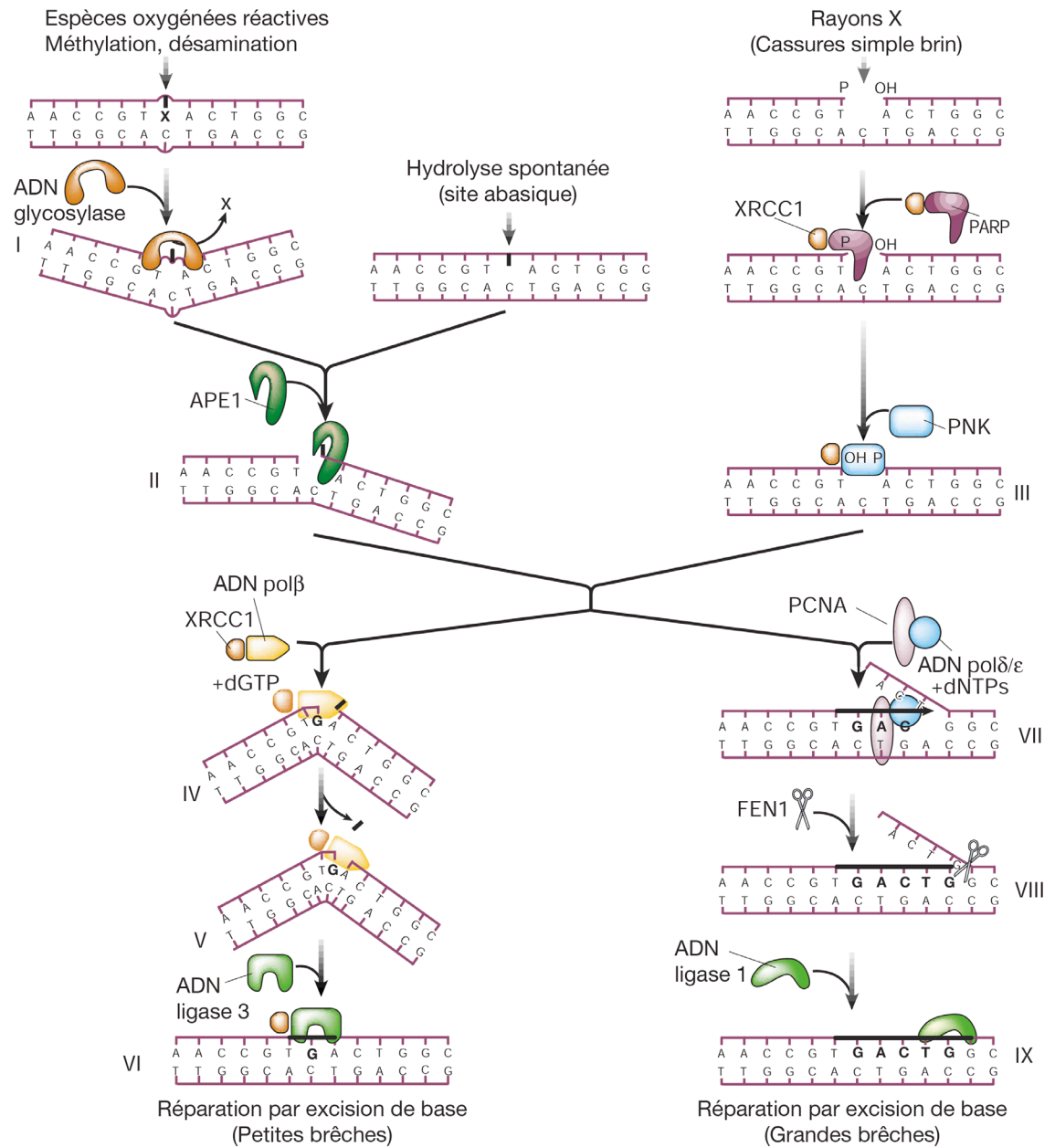
La tolérance aux lésions est un processus permettant aux cellules de répliquer de l'ADN endommagé. Ce processus commence lorsqu'une fourche de réplication arrive au niveau d'une région endommagée. Il y a alors arrêt de la réplication de manière à permettre l'échange de la polymérase répliquative qui copie l'ADN avec une grande fidélité par une polymérase spécialisée dans le franchissement des lésions (87). Celles-ci possèdent un site actif plus vaste que les polymérases répliquatives, ce qui leur permet d'accommoder des canevas d'ADN modifiés et donc de copier de l'ADN endommagé (88). En revanche, elles ne

possèdent pas d'activité correctrice, et sont donc plus susceptibles d'introduire des mutations dans l'ADN que les polymérases réplcatives. Les organites de plantes semblent dépourvus de systèmes de tolérance aux lésions. En effet, aucun gène codant pour une polymérase spécialisée dans le franchissement des lésions et possédant un signal de localisation vers les organites n'a été détecté dans les génomes de plantes (89).

La réparation de l'ADN par excision de base

La réparation de l'ADN par excision de base est utilisée lorsqu'une base azotée subit une modification chimique (Figure 1-3). Chez les mammifères, il existe deux voies de réparation par excision de base : la voie par remplissage de petites brèches (voie principale) et la voie par remplissage de grandes brèches (voie marginale). La première voie commence par la reconnaissance de la base azotée modifiée par une ADN glycosylase (90). Celle-ci hydrolyse le lien covalent entre la base azotée modifiée et le sucre de manière à former un site abasique. Une cellule peut posséder plusieurs types d'ADN glycosylases qui reconnaissent chacune un spectre défini de modifications chimiques. Une fois la base azotée modifiée clivée, le site abasique va être reconnu par une endonucléase qui va inciser l'ADN au niveau du site abasique. Une polymérase pourra ensuite éliminer le site abasique grâce à son activité lyase avant de procéder à la resynthèse d'ADN. La brèche sera finalement scellée grâce à l'action d'une ADN ligase. Contrairement à la voie de réparation des petites brèches qui ne nécessite que l'excision du nucléotide fautif, la voie de réparation des grandes brèches nécessite quant à elle l'excision d'une dizaine de nucléotides. Récemment, deux études ont montré l'existence d'un système d'excision de base par remplissage de petites brèches dans les organites de plantes. En effet, des homologues des composantes de la machinerie d'excision de base ont été identifiés dans les chloroplastes d'*Arabidopsis thaliana* (91). De plus, une activité d'excision de bases oxydées a pu être mise en évidence dans des extraits protéiques de chloroplastes d'*Arabidopsis thaliana* (91). Dans une autre étude, des substrats d'ADN contenant des bases uraciles ont été importés dans des mitochondries purifiées d'*Arabidopsis thaliana* (92). La réparation de ces substrats a pu être observée, montrant ainsi que les mitochondries possèdent un système d'excision de base. Ce système est tributaire de l'action d'une glycosylase d'uracile et d'une endonucléase de site abasique (92).

Figure 1-3. Mécanisme de réparation de l'ADN par excision de bases chez les mammifères. Il existe trois voies d'entrée à la réparation de l'ADN par excision de bases. Tout d'abord, une ADN glycosylase peut reconnaître une base modifiée et cliver cette base de manière à former un site abasique (I gauche). Celui-ci peut également être formé suite à l'hydrolyse spontanée du lien sucre-base (I centre). Aussi, des rayons ionisants peuvent entraîner une cassure monocaténaire qui sera reconnue par un complexe entre la Poly(ADP-ribose) polymérase (PARP) et la kinase de polynucléotides (PNK) (I droite). La voie de remplissage de petites brèches commence par la reconnaissance du site abasique par l'endonucléase APE1 qui va inciser l'ADN à proximité du site abasique (II). Une ADN polymérase β va ensuite relâcher le nucléotide abasique grâce à son activité lyase (IV) et combler la brèche par synthèse d'ADN (V). Finalement, un complexe formé par XRCC1 et l'ADN ligase 3 va sceller l'ADN (VI). Dans la voie de réparation par remplissage de grandes brèches, un complexe formé par les ADN polymérases δ/ϵ et par PCNA va synthétiser de l'ADN dans la brèche et ainsi déplacer deux à dix nucléotides (VII). L'endonucléase FEN1 va supprimer ces nucléotides déplacés (VIII) après quoi l'ADN ligase 1 va sceller l'ADN (IX). Cette figure est adaptée de (81) avec la permission de Macmillan Publishers Ltd (copyright 2001).



La réparation de l'ADN par excision de nucléotides

Le système d'excision de nucléotides permet la détection et la réparation d'un grand nombre de modifications survenant à l'ADN : modification chimique de nucléotides, formation de dimères de pyrimidine, pontage interbrin ou liaison covalente d'une protéine à l'ADN (Figure 1-5). La flexibilité de ce système de réparation est dû au fait qu'il ne reconnaît pas directement la modification de l'ADN, mais plutôt la déformation de la double hélice d'ADN découlant de sa modification (93). Il existe deux voies de détection des nucléotides modifiés. Une première, dite couplée à la transcription, est activée lorsqu'un dommage de l'ADN situé dans un gène provoque l'arrêt de la machinerie de transcription. Une deuxième voie, dite globale, agit dans tous les autres cas. Ces voies ne diffèrent que par la manière dont la modification de l'ADN est reconnue. Dans le cas de la réparation couplée à la transcription chez les mammifères, la modification de l'ADN provoque l'arrêt de l'ARN polymérase et le recrutement de protéines de réparation dont TFIIH. Dans le cas de la réparation globale chez les mammifères, la modification de l'ADN est détectée par le complexe XPC-HR23B qui recrute alors des protéines de réparation dont TFIIH. Cette dernière possède une activité hélicase qui va lui permettre de dénaturer localement l'ADN au niveau du site endommagé. Il y aura alors excision, par une endonucléase, de 24 à 32 nucléotides du brin contenant le nucléotide endommagé, resynthèse d'ADN par une polymérase à ADN et scellage du site par une ADN ligase. Il ne semble pas y avoir de système d'excision de nucléotides chez les organites de plantes.

La correction des mésappariements dans l'ADN

Des erreurs lors de la réplication de l'ADN peuvent entraîner des mésappariements dans l'ADN ou l'insertion de boucles due au décalage de la synthèse d'ADN sur le brin matrice. Ces changements dans l'ADN peuvent être détectés et réparés par un système de correction des mésappariements (Pour revue, se référer à (94); Figure 1-5). Ce système reconnaît le mésappariement et identifie le brin parental contenant l'information génétique à conserver. Le brin néosynthétisé est alors dégradé grâce à la collaboration entre une protéine hélicase et une protéine exonucléase. L'ADN dégradé sera resynthétisé par une polymérase à ADN après quoi une ligase pourra sceller l'ADN. Chez les bactéries, les

protéines MutS, MutL et MutH sont impliquées dans la correction des mésappariements d'ADN. Chez *Arabidopsis thaliana*, un homologue de MutS, Msh1, a été détecté dans la mitochondrie (45,95). Cependant, Msh1 semble plutôt impliqué dans le contrôle de la recombinaison que dans la correction des mésappariements dans l'ADN. En effet, la suppression de l'expression du gène codant pour Msh1 n'entraîne pas d'augmentation dans la quantité de mutations détectées dans le génome mitochondrial (45). En somme, il ne semble pas y avoir de système de correction des mésappariements dans l'ADN des organites de plantes.

La réparation des cassures bicaténares de l'ADN

Les deux principales voies de réparation des cassures bicaténares sont la recombinaison homologue et la jonction d'extrémités non homologues (Figure 1-6). Ces voies se distinguent par leur fidélité de réparation ainsi que par les composantes protéiques impliquées dans la réparation. Bien que la plupart des organismes possèdent ces deux voies de réparation, certains organismes vont favoriser une voie plutôt qu'une autre pour réparer les cassures bicaténares d'ADN. Ainsi, la recombinaison homologue est utilisée préférentiellement chez les bactéries et les levures tandis que la jonction d'extrémités non homologues semble être préférentiellement utilisée chez les plantes et chez les vertébrés (96). Le choix d'une voie de réparation est également fonction du cycle cellulaire et de la nature des cassures bicaténares (96). Récemment, plusieurs autres voies ont été découvertes. Ces nouvelles voies semblent prendre le relais lorsque l'une ou l'autre des voies principales n'est pas disponible. Cette hypothèse concorde avec des expériences suggérant une complémentarité fonctionnelle des voies de réparation des cassures bicaténares (96).

La réparation par jonction d'extrémités non homologues

La jonction d'extrémités non homologues permet de joindre deux extrémités d'ADN sans égard à la séquence présente à chacune des extrémités. Contrairement à la recombinaison homologue, cette voie ne nécessite pas l'utilisation d'un ADN de séquence similaire à celui endommagé. Théoriquement, cette voie peut permettre la jonction des extrémités d'ADN sans perte d'information. En pratique, cependant, les dommages dus à des composés chimiques ou à du rayonnement électromagnétique de haute énergie rendent

Figure 1-4. Mécanisme de réparation de l'ADN par excision de nucléotides chez les mammifères. La voie de réparation globale commence par la détection d'une distorsion de l'ADN par le complexe XPC-HR23B (I gauche). La voie couplée à la transcription commence quant à elle par l'arrêt de l'ARN polymérase au niveau d'un dommage à l'ADN et le relargage de l'ARN polymérase par les protéines CSA et CSB (I droite). Dans les deux cas, il y aura ouverture d'environ trente nucléotides d'ADN au niveau du site endommagé par la protéine TFIIH (II). La protéine RPA permettra de stabiliser le brin non endommagé (III) tandis que le brin endommagé sera excisé par les endonucléases XPG et ERCC1/XPF (IV). La machinerie de réplication de l'ADN pourra finalement remplir la brèche et sceller l'ADN (V). Cette figure est adaptée de (81) avec la permission de Macmillan Publishers Ltd (copyright 2001).

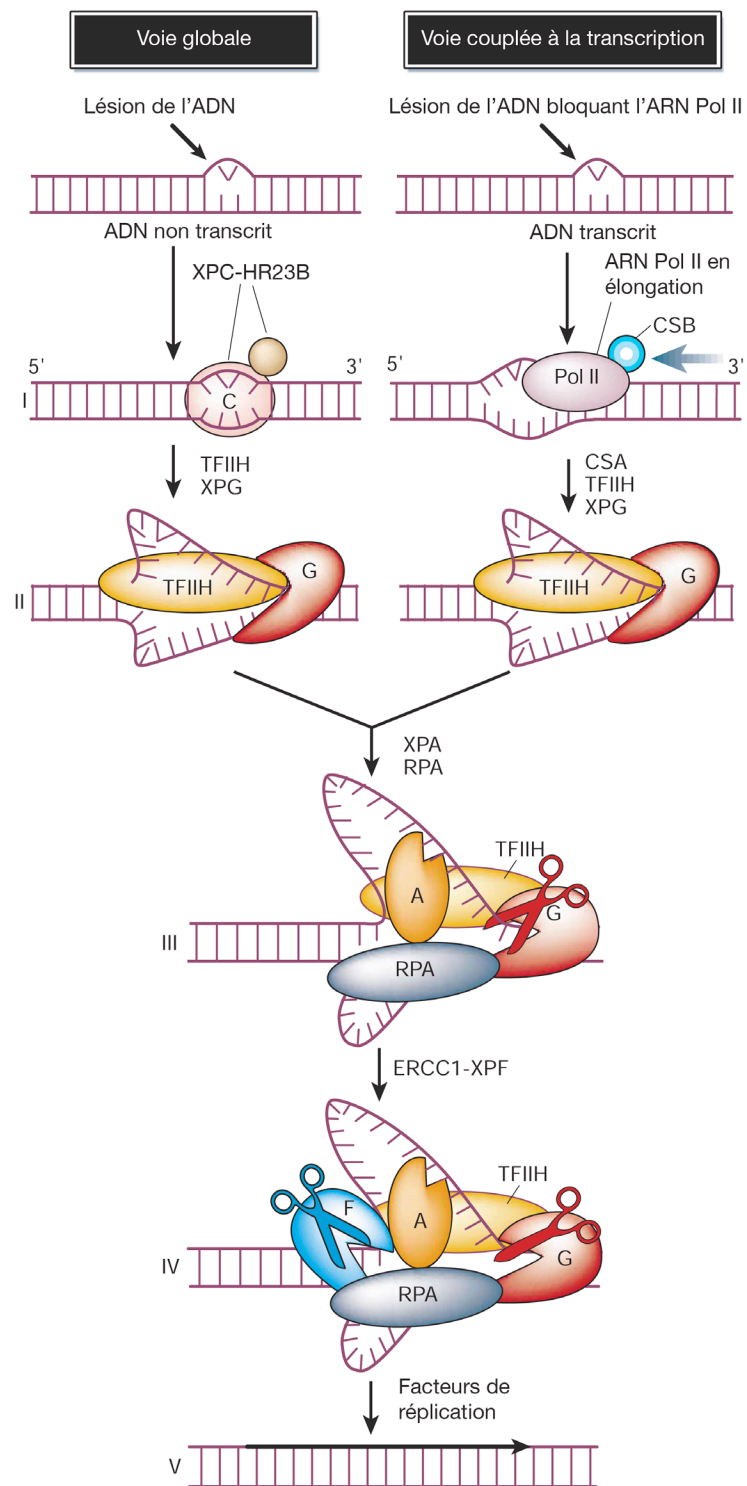


Figure 1-5. Mécanisme de correction des mésappariements dans l'ADN chez les mammifères. Les protéines MSH2/6 reconnaissent la présence de mésappariements dans l'ADN tandis que les protéines MSH2/3 reconnaissent des boucles formées par insertion ou délétion d'ADN durant la synthèse. La reconnaissance du brin parental se fait en communiquant avec la machinerie de réplication de l'ADN. Une fois le complexe MSH2/6-ADN formé (I), il y aura dégradation de l'ADN néosynthétisé par une exonucléase 3'→5' (II). En revanche, une fois que le complexe MSH2/3-ADN est formé (II), il y a plutôt dégradation de l'ADN néosynthétisé par une exonucléase 5'→3' (III). Une fois l'ADN digéré, il pourra y avoir reprise de la synthèse d'ADN. Cette figure est adaptée de (81) avec la permission de Macmillan Publishers Ltd (copyright 2001).

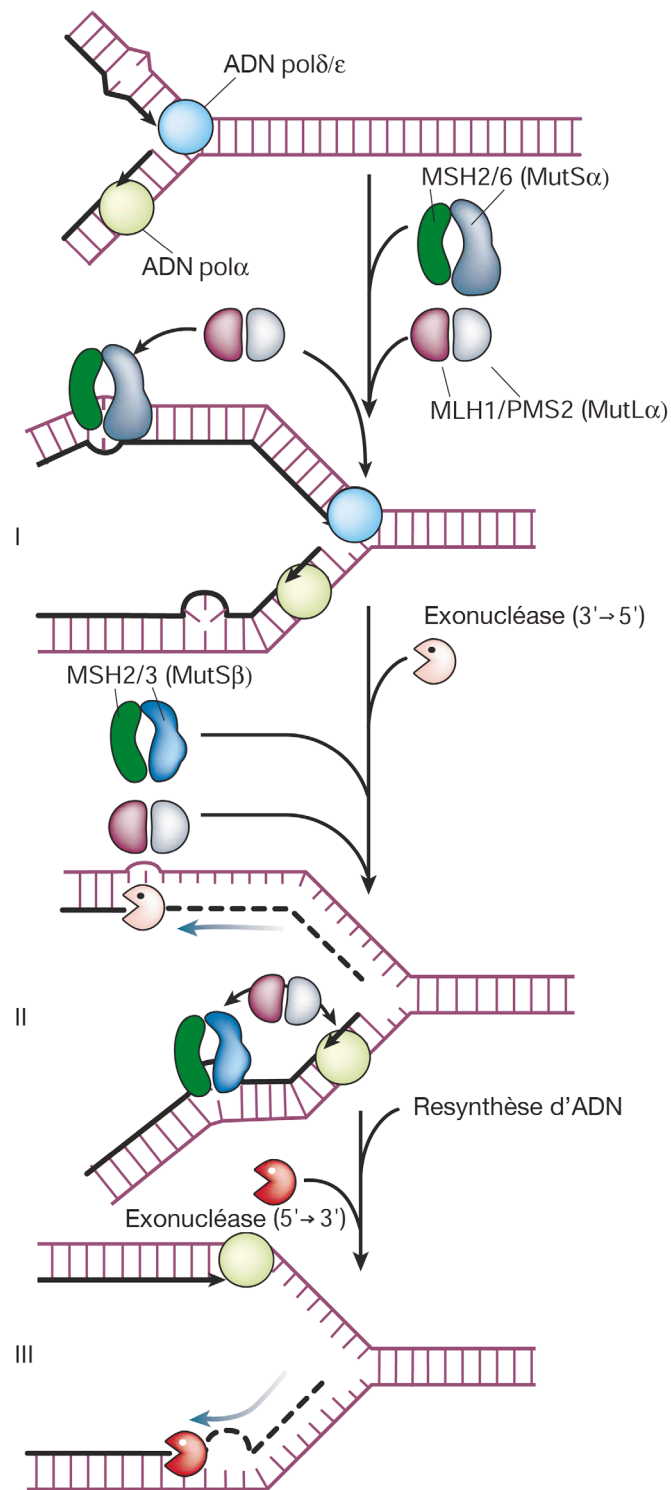
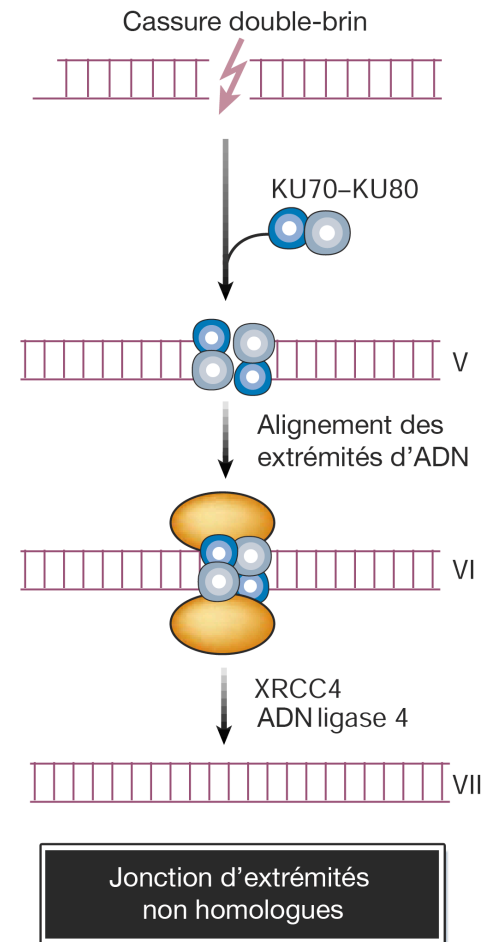
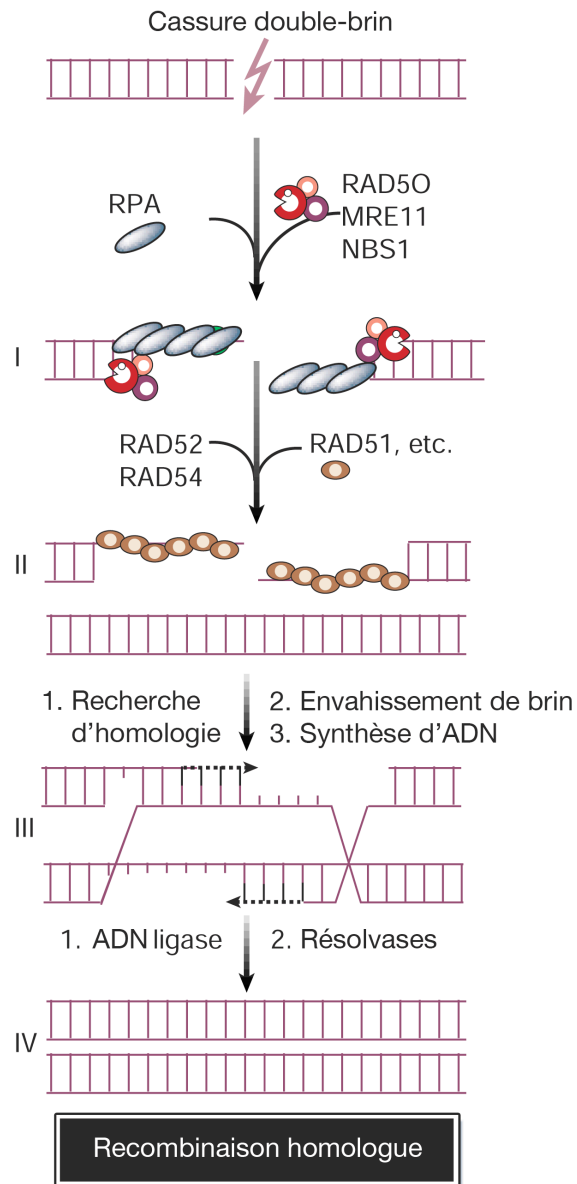


Figure 1-6. Mécanismes de réparation des cassures bicaténaire de l'ADN par la voie de la recombinaison homologue ou par la voie de la jonction d'extrémités non-homologues chez les mammifères. Après formation d'une cassure bicaténaire, il peut y avoir résection de l'ADN grâce à l'activité exonucléase du complexe RAD50/MRE11/NBS1 et protection de l'ADN monocaténaire par la protéine RPA (I). Cette protéine facilite l'assemblage d'un filament de protéine RAD51 sur l'ADN monocaténaire (II). RAD51 va ensuite rechercher un ADN bicaténaire homologue à l'ADN monocaténaire auquel il est lié et catalyser un échange de brins qui donne origine à des structures d'ADN cruciformes appelées jonctions de Holliday. L'ADN homologue va alors servir de matrice pour réparer le brin cassé par synthèse d'ADN (III). Des ADN résolvases vont finalement permettre de résoudre les jonctions de Holliday et de séparer les brins (IV). Dans la voie de la jonction d'extrémités non-homologues, les extrémités cassées sont reconnues par les protéines Ku70/80 (V). Ces protéines protègent les extrémités de l'ADN et les empêchent de se séparer. En conjonction avec les protéines ADN-PKCS, les protéines Ku70/80 vont aligner les extrémités d'ADN (VI) et l'ADN sera alors scellé par les protéines XRCC4-ligase 4. Cette figure est adaptée de (81) avec la permission de Macmillan Publishers Ltd (copyright 2001).



fréquemment les extrémités non joignables (97). Pour pallier ce problème, les extrémités de l'ADN sont digérées avant leur jonction, ce qui entraîne une perte de nucléotides au site de cassure de l'ADN. Au cours de la réparation par jonction d'extrémités non homologues, les cassures bicaténares vont être reconnues par les protéines hétérodimériques Ku70/80 qui vont se fixer sur chacune des extrémités bicaténares de l'ADN et en empêcher la digestion ou l'éloignement. L'ADN sera par la suite pris en charge par des exonucléases qui vont permettre l'élimination des extrémités non joignables. Des ligases pourront ensuite catalyser la jonction des extrémités (pour revue, se référer à (97) et (98)). Plusieurs évidences suggèrent l'absence d'un système de réparation de l'ADN par jonction d'extrémités non homologues dans les plastides. En effet, aucun gène codant pour une protéine impliquée dans ce type de réparation et dirigée vers le chloroplaste n'a pu être identifié chez *Chlamydomonas reinhardtii* (99). De plus, la transformation de chloroplastes avec de l'ADN montre que l'insertion dans le génome s'effectue via différents mécanismes mais pas via une jonction d'extrémités non homologues (99,100).

Les voies alternatives de jonctions d'extrémités non homologues

Récemment, plusieurs études ont montré l'existence de voies alternatives de jonction d'extrémités non homologues. Celles-ci utilisent des composantes protéiques différentes que la voie classique. En outre, des très courtes séquences répétées, ou *microhomologies*, sont fréquemment utilisées pour permettre la jonction des molécules d'ADN (101,102). Une étude génétique chez la levure a permis d'identifier les protéines impliquées dans cette voie de réparation (103). Bien que des réarrangements de l'ADN impliquant des très courtes séquences répétées aient été observés dans les organites de plantes (57,104-106), on ignore si ceux-ci sont obtenus suite à la réparation de l'ADN par des voies alternatives de la jonction d'extrémités non homologues.

La réparation par recombinaison homologue

La recombinaison homologue permet de réparer les cassures bicaténares sans perte d'information. Ce type de réparation nécessite la présence, dans le génome, d'une séquence identique ou tout au moins fortement similaire à celle ayant été endommagée. Le processus de recombinaison est très sensible à la longueur ainsi qu'au degré de similarité des séquences

homologues. Il nécessite en effet que les séquences homologues aient une taille supérieure à 50 paires de bases chez *E. coli* (107) et supérieure à 300 chez les mammifères (108). Les principales étapes de ce processus sont : la formation d'une extension 3' d'ADN monocaténaire par l'action d'exonucléases 5'-3', sa stabilisation par des protéines liant l'ADN monocaténaire et le recrutement d'une protéine recombinase. Celle-ci va rechercher un ADN homologue et catalyser l'invasion de l'ADN bicaténaire par l'ADN monocaténaire. Ceci donne lieu à un échange de brin ainsi qu'à la formation d'une jonction de Holliday. Il pourra alors y avoir synthèse d'ADN à partir de l'extrémité 3' de l'ADN brisé en utilisant la séquence homologue comme matrice. À partir de ce moment, plusieurs dénouements sont possibles : (i) résolution des jonctions de Holliday avec *crossover*, (ii) résolution des jonctions de Holliday sans *crossover* et (iii) libération de l'ADN par une hélicase. Dans ce cas, l'ADN néosynthétisé va pouvoir s'apparier avec l'autre extrémité de la cassure bicaténaire. La recombinaison homologue est active dans les organites de plantes, où elle joue de nombreux rôles allant de l'organisation du génome à l'élimination des mutations. Des homologues de la protéine RecA de bactérie ont ainsi été identifiés et caractérisés dans les chloroplastes (47,109,110) ainsi que dans les mitochondries (47,48,111) de plusieurs plantes. De plus, des intermédiaires de recombinaison ainsi qu'une activité d'échange de brins a pu être mise en évidence dans des extraits mitochondriaux de haricots (112).

La réparation par appariement d'ADN monocaténaire

La réparation par appariement d'ADN monocaténaire (*Single Strand Annealing (SSA)*) est parfois considérée comme un cas particulier de recombinaison homologue. Ce type de réparation ne peut survenir que lorsqu'une cassure bicaténaire survient entre des séquences répétées. L'ADN est digéré par des exonucléases à partir du site de cassure de manière à former une extension 3' d'ADN monocaténaire et d'exposer les séquences répétées qui pourront être appariées grâce à l'action de la protéine Rad52 (113). L'ADN monocaténaire protubérant est par la suite digéré par des nucléases. Finalement, l'ADN bicaténaire est reconstitué par synthèse d'ADN et scellé par des ADN ligases. De par son mode de fonctionnement, ce type de réparation conduit à la délétion de l'ADN situé entre les séquences répétées ainsi qu'à la délétion d'une des séquences répétées. Chez les mammifères, ce mécanisme se fait de manière Rad52-dépendant et Rad51-indépendant. Un

tel mécanisme semble être fonctionnel dans les chloroplastes. En effet, des chloroplastes de l'algue verte *Chlamydomonas reinhardtii*, transformés avec de l'ADN, incorporent cet ADN par un mécanisme semblable à la voie de réparation par appariement d'ADN monocaténaire (99).

La réplication induite par les cassures

La réplication induite par les cassures (*Break Induced Replication (BIR)*; pour revue, se référer à (114,115)) est également considérée comme un cas particulier de recombinaison homologue et est principalement utilisée pour réparer les cassures bicaténares d'ADN n'ayant qu'une seule extrémité libre qui sont notamment produites lors de l'effondrement d'une fourche de réplication. Au cours de ce processus, l'extrémité 3' de la molécule endommagée va envahir une molécule d'ADN homologue de manière à reformer une fourche de réplication. Il existe une variante de ce processus, appelé réplication induite par les cassures dépendantes de très courtes séquences répétées (*Microhomology Mediated Break Induced Replication*), qui ne nécessite que la présence de très courtes séquences répétées pour permettre le redémarrage de la réplication de l'ADN (116-118). Ces voies peuvent mener à la formation de réarrangements complexes étant donné qu'après l'envahissement d'une molécule homologue et le redémarrage de la fourche de réplication, l'ADN endommagé peut se dissocier de sa matrice de réparation de manière à pouvoir envahir une nouvelle séquence d'ADN (119).

Coordination des différentes voies de réparation de l'ADN

L'analyse génétique des différentes voies de réparation de l'ADN a conduit à l'identification de certaines de leurs composantes protéiques. Avec la découverte de nouvelles voies de réparation de l'ADN, on s'aperçoit que plusieurs des protéines agissant dans une voie peuvent également accomplir leur fonction dans le cadre d'une autre voie. Au niveau du site de cassure, plusieurs protéines peuvent ainsi entrer en compétition pour pouvoir lier les extrémités d'ADN. Les protéines liées à l'ADN vont ainsi favoriser le recrutement de leurs partenaires d'interaction et par là même aiguiller la réparation des cassures d'ADN vers une certaine voie (Pour revue, se référer à (101)).

1.2 Les protéines liant les acides nucléiques monocaténares

Les acides nucléiques monocaténares interviennent dans le cadre de processus biologiques comme la réplication de l'ADN, la recombinaison de l'ADN, la réparation de l'ADN, la maturation des ARN, la traduction des ARN et la formation de complexes nucléoprotéiques (120). Sous leur forme monocaténaire, les acides nucléiques (*i*) sont plus vulnérables à la dégradation par des nucléases, (*ii*) sont plus facilement victimes de modifications chimiques comme la dépurination ou la désamination, (*iii*) ont un potentiel recombino-génique plus élevé et (*iv*) peuvent former des structures secondaires ou s'apparier avec un brin complémentaire (120). Les protéines liant les acides nucléiques monocaténares ont probablement évolué de manière à accomplir leurs fonctions tout en limitant ces avenues non souhaitables.

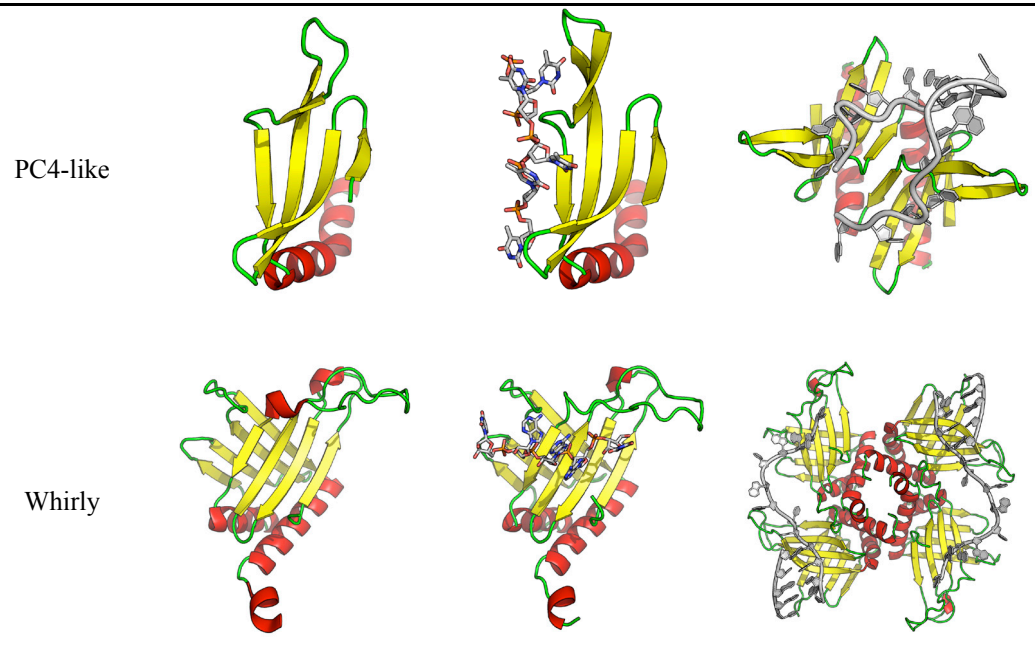
Nous allons voir dans cette section la structure des principaux domaines de liaison aux acides nucléiques monocaténares. Aussi, nous ferons ressortir les principes généraux concernant la liaison des protéines aux acides nucléiques monocaténares grâce à l'étude de quelques exemples.

1.2.1 Structure des domaines de liaison aux acides nucléiques monocaténares

L'analyse de la structure des protéines liant les acides nucléiques monocaténares mène à plusieurs constats. Premièrement, les feuillets β sont fréquemment impliqués dans la liaison des acides nucléiques monocaténares. Ceci contraste avec la liaison des acides nucléiques bicaténares où des hélices α sont généralement utilisées. Deuxièmement, seul un petit nombre de motifs structuraux sont impliqués dans la liaison des acides nucléiques monocaténares (Tableau 1-III). Ceux-ci se classent en différentes familles selon l'agencement tridimensionnel de leurs éléments de structure secondaire. Troisièmement, chaque famille regroupe des protéines qui peuvent posséder des séquences, des activités biochimiques et des fonctions différentes. Cette diversité est généralement associée à l'apparition très précoce, au cours de l'évolution, de ces motifs structuraux. Ainsi, les gènes codant pour les protéines comportant ces motifs ont eu le temps d'être dupliqués et les différentes copies ont pu évoluer indépendamment les unes des autres.

Tableau 1-III. Structure des domaines de liaison aux acides nucléiques monocaténares

Domaine	Organisation du domaine (forme non liée aux acides nucléiques)	Organisation du domaine (forme liée aux acides nucléiques)	Organisation de la protéine (forme liée aux acides nucléiques)
OB-fold	A ribbon diagram of an OB-fold domain, showing a beta-sheet (yellow) and an alpha-helix (red) connected by loops (green).	A ribbon diagram of an OB-fold domain bound to a single-stranded nucleic acid (blue and red spheres), showing the interaction between the beta-sheet and the nucleic acid.	A ribbon diagram of a protein with multiple OB-fold domains bound to a single-stranded nucleic acid, showing the overall architecture of the protein-nucleic acid complex.
RRM	A ribbon diagram of an RRM domain, showing a beta-sheet (yellow) and an alpha-helix (red) connected by loops (green).	A ribbon diagram of an RRM domain bound to a single-stranded nucleic acid (blue and red spheres), showing the interaction between the beta-sheet and the nucleic acid.	A ribbon diagram of a protein with multiple RRM domains bound to a single-stranded nucleic acid, showing the overall architecture of the protein-nucleic acid complex.
KH	A ribbon diagram of a KH domain, showing a beta-sheet (yellow) and an alpha-helix (red) connected by loops (green).	A ribbon diagram of a KH domain bound to a single-stranded nucleic acid (blue and red spheres), showing the interaction between the beta-sheet and the nucleic acid.	A ribbon diagram of a protein with multiple KH domains bound to a single-stranded nucleic acid, showing the overall architecture of the protein-nucleic acid complex.
Sm-fold	A ribbon diagram of a Sm-fold domain, showing a beta-sheet (yellow) and an alpha-helix (red) connected by loops (green).	A ribbon diagram of a Sm-fold domain bound to a single-stranded nucleic acid (blue and red spheres), showing the interaction between the beta-sheet and the nucleic acid.	A ribbon diagram of a protein with multiple Sm-fold domains bound to a single-stranded nucleic acid, showing the overall architecture of the protein-nucleic acid complex.



Les modèles suivants ont été utilisés pour obtenir ces images. OB-fold, Structures cristallines de Hs-RPA (code pdb 1FGU) et de Hs-RPA-d(C8) (code pdb 1JMC); RRM, Structures cristallines de Hs-hnRNPA1 (code pdb 1L3K) et de Hs-hnRNPA1-d(TTAGGGTTAGGG) (code pdb 2UP1); KH, Structures cristallines de Ec-ERA (code pdb 1EGA) et de Aa-ERA-r(AUCACCUCCUUA) (code pdb 3IEV); Sm-fold, Structures cristallines de Sa-Hfq (code pdb 1KQ1) et de Sa-Hfq-r(AUUUUUG) (code pdb 1KQ2); PC4-like, Structures cristallines de Hs-PC4 (code pdb 1PCF) et de Hs-PC4-d(T19G) (code pdb 2C62); Whirly, Structures cristallines de Tb-MRP1/2 (code pdb 2GIA) et de Tb-MRP1/2-ARN guide (code pdb 2GJE). Les hélices α sont colorées en rouge, les feuillets β sont colorés en jaune et les boucles sont colorées en vert. Les carbones des acides nucléiques monocaténares sont colorés en gris.

1.2.1.1 Le domaine OB-fold

Le OB (Oligonucleotide/oligosaccharide Binding)-fold est un domaine d'une longueur variant de 70 à 150 acides aminés (121). On le définit minimalement comme l'assemblage en baril β d'un feuillet antiparallèle de cinq brins β généralement fermé à une extrémité par une hélice α . Il n'existe pas de séquence consensus pour les protéines à domaine OB-fold, ce qui rend leur détection difficile à partir des séquences d'acides aminés (120,122). Le domaine OB-fold semble avoir émergé très tôt durant l'évolution, étant donné qu'on le retrouve chez les archées, les procaryotes et les eucaryotes (123). Bien que la majorité des protéines de la famille OB-fold reconnaissent l'ADN ou l'ARN monocaténaire, certaines reconnaissent plutôt des sucres ou des peptides (121). Le domaine OB-fold des ligases d'ADN des organismes eucaryotes reconnaissent quant à elles de l'ADN bicaténaire (124). Les protéines à domaine OB-fold sont principalement impliquées dans la réplication, la recombinaison, la réparation, la transcription, la réponse au stress hypothermique et la maintenance des télomères (121). Certaines de ces protéines, comme SSB ou RPA, se lient aux acides nucléiques monocaténaires sans spécificité de séquence, tandis que d'autres protéines comme Pot1, Cdc13 ou TEBP possèdent une très forte spécificité de séquence. Certaines protéines possèdent plusieurs domaines OB-fold disposés en tandem et peuvent adopter une panoplie de structures quaternaires (121).

Une recherche dans la banque de donnée PDB révèle qu'il existe plusieurs centaines de structures de protéines à domaine OB-fold dont une cinquantaine présentent la protéine en complexe avec des acides nucléiques monocaténaires. L'analyse de ces structures nous apprend que l'interaction avec les acides nucléiques monocaténaires s'effectue au niveau d'une cavité formée par les feuillets β . Les bases azotées interagissent généralement avec des acides aminés hydrophobes ou aromatiques localisés dans la cavité tandis que les groupements phosphates sont exposés au solvant. La cavité est bordée par deux boucles appelées L12 et L45 qui contiennent des acides aminés basiques ou aromatiques et qui peuvent enclaver les acides nucléiques (122).

1.2.1.2 Le domaine RRM

Le RRM (*RNA Recognition Motif*) est un domaine d'environ 90 acides aminés (125). Il est minimalement composé de deux hélices α plaquées contre un plancher β antiparallèle constitué de quatre brins β et possède une topologie de type $\beta\alpha\beta\beta\alpha\beta$. Il contient deux motifs, RNP1 et RNP2, dont la séquence est relativement bien conservée. Ces motifs, qui ont permis l'identification de nouvelles protéines à domaine RRM, sont localisés sur les deux brins β situés au centre du plancher β (125). Tout comme le domaine OB-fold, le domaine RRM est très ancien. Des gènes codant pour des protéines à domaine RRM sont en effet présents chez tous les domaines du vivant et ils sont particulièrement abondants dans le génome des eucaryotes (125). Chez l'humain, on estime ainsi qu'environ un pour cent des gènes codent pour des protéines contenant au moins un domaine RRM (126). La majorité des protéines à domaine RRM reconnaissent l'ARN monocaténaire de manière séquence spécifique. Certaines reconnaissent toutefois l'ADN monocaténaire ou des protéines (127). Les protéines à domaine RRM sont principalement impliquées dans la maturation des ARN, le transport des ARN et la régulation de la stabilité des ARN (127). Elles possèdent très souvent plusieurs domaines RRM disposés en tandem (126), ce qui leur permet d'accroître leur spécificité de reconnaissance.

On connaît la structure d'une vingtaine de protéines à domaine RRM en complexe avec des acides nucléiques (126). Leur analyse montre que le plancher β constitue le site principal de liaison aux acides nucléiques. Celui-ci peut accommoder de trois à quatre nucléotides (125). Trois acides aminés aromatiques appartenant aux deux motifs RNP permettent de stabiliser deux nucléotides sur le plancher β de manière séquence indépendante. La liaison spécifique de ces nucléotides est obtenue grâce à des acides aminés situés en dehors des motifs RNP qui interagissent avec les extrémités fonctionnelles des bases azotées. Cette stratégie de liaison confère une grande adaptabilité au domaine RRM car un changement de spécificité de séquence peut se faire par la seule modification de quelques acides aminés.

1.2.1.3 Le domaine KH

Le KH (*K Homology*) est un domaine de 70 acides aminés (128) découvert la première fois chez la protéine hnRNP K (129). Ce domaine, constitué de trois hélices α et de trois brins β , possède généralement un motif dont la séquence consensus est VIGXXGXXI. Il existe deux types de domaines KH qui diffèrent de par l'arrangement tridimensionnel de leurs éléments de structure secondaire. Le type I possède une topologie $\beta\alpha\alpha\beta\beta\alpha$ avec des feuillets β antiparallèles et est généralement retrouvé en copies multiples chez les protéines eucaryotes, tandis que le type II possède une topologie $\alpha\beta\beta\alpha\alpha\beta$ avec des feuillets β mixtes et est retrouvé en copie simple chez les protéines procaryotes. En plus des procaryotes et des eucaryotes, des protéines à domaine KH sont également présents chez les archées, ce qui suggère que le domaine KH est apparu très tôt au cours de l'évolution (130). Les protéines à domaines KH reconnaissent l'ARN ou l'ADN monocaténaire et sont souvent impliquées dans la régulation de la transcription, l'épissage des ARN ou la régulation de la traduction des ARN. Elles utilisent généralement une paire de domaines KH en tandem (120).

La surface de liaison aux acides nucléiques du domaine KH est constituée de deux hélices α , de la boucle GXXG, et de l'extrémité d'un brin β . Cette surface permet d'accommoder quatre nucléotides (128). Il y a peu d'empilement aromatique entre les acides aminés du domaine KH et les nucléotides (126). Des forces hydrophobes permettent cependant de stabiliser les nucléotides. Finalement, la comparaison de la structure de protéines à domaine KH en forme libre et en forme liée à l'ARN ou à l'ADN monocaténaire montre que le processus de liaison n'entraîne pas de changement dans la conformation de la protéine.

1.2.1.4 Le domaine Sm-fold

Le Sm-fold est un domaine d'environ 60 acides aminés. Il est composé d'une hélice α et de cinq brins β qui forment un plancher antiparallèle (120). Le premier et le dernier brin β sont courts, tandis que les trois autres brins β sont longs et coudés. Les protéines à domaine Sm-fold sont présentes chez les archées, les procaryotes et les eucaryotes. Les protéines à domaine Sm-fold des procaryotes s'assemblent en homohexamères de manière à former un anneau de sept nanomètres de diamètre. Chez les eucaryotes, ces protéines

s'assemblent en hétéroheptamères en forme d'anneau (131). Cet assemblage est facilité par des interactions entre les extrémités des planchers β voisins et grâce à l'interaction entre l'hélice α d'une protéine et le plancher β d'une protéine adjacente. Les protéines à domaine Sm-fold lient l'ARN avec une préférence pour les séquences riches en uridines et participent à la biogénèse des petites ribonucléoprotéines nucléaires ainsi qu'à la maturation et à la dégradation de l'ARN (131).

La structure de quelques protéines à domaine Sm-fold en complexe avec de l'ARN a été élucidée. L'ARN se lie à la cavité centrale de l'anneau grâce à des acides aminés positionnés sur des boucles pointant vers le centre de l'anneau (120). Ces boucles forment des poches dans lesquelles une base peut prendre place et interagir avec deux résidus aromatiques et avec des acides aminés basiques. Une asparagine située à proximité de la poche favorise la liaison avec une base uracile.

1.2.1.5 Le domaine PC4

Le domaine PC4 (*Positive Cofactor 4*) mesure environ 60 acides aminés. Il est composé d'une hélice α et de cinq brins β qui forment un plancher antiparallèle. Les protéines à domaine PC4 forment minimalement des dimères qui sont stabilisés grâce à l'interaction entre l'hélice α d'une protéine et le plancher β de l'autre protéine. On ne les retrouve que chez les eucaryotes. Chez l'humain, la protéine PC4 intervient au niveau de la régulation de la transcription de nombreux gènes (132), de l'organisation de la chromatine et de la réparation des cassures bicaténares. Elle a une longueur de 127 acides aminés et comprend deux régions. La région C-terminale comprend un domaine de liaison à l'ADN monocaténaire (133,134). Cette région est également le site d'interaction avec des activateurs transcriptionnels tels que VP16 et p53 (135-137). La région N-terminale ne possède pas de structure tridimensionnelle stable mais module l'interaction de PC4 avec l'ADN ou les activateurs transcriptionnels (138). Cette région est phosphorylée par la protéine CK2, ce qui entraîne un changement dans la capacité de PC4 à lier l'ADN et les activateurs transcriptionnels (139).

L'interaction avec l'ADN monocaténaire s'effectue au niveau de la surface concave formée par les planchers β . Des acides aminés basiques et aromatiques stabilisent l'ADN

dans une conformation en épingle à cheveux. La protéine PC4 forme un homodimère grâce à des interactions entre les planchers β de protéines adjacentes. Sous cette forme, PC4 peut lier les deux brins d'une molécule d'ADN dénaturée (134).

1.2.1.6 Le domaine Whirly

Le domaine Whirly a une longueur d'environ 160 acides aminés. Chaque domaine est constitué de deux planchers β antiparallèles constitués chacun de quatre brins β apposés contre deux hélices α . Une troisième hélice α , optionnelle, interagit avec l'hélice α d'une sous-unité adjacente. Il existe plusieurs protéines à domaine Whirly : (i) les protéines Whirly présentes chez les plantes (140), (ii) les protéines MRP1 et MRP2 présentes chez les trypanosomatides (141), (iii) les protéines Pur α présentes chez les eucaryotes (142), (iv) les protéines SsbG présentes chez les actinomycètes (143) et (v) deux protéines hypothétiques présentes principalement chez les cyanobactéries (Tableau 1-IV). Les protéines St-WHY1 et MRP1/2 forment des tétramères avec une symétrie de type C4 ou pseudo-C4 (140,141). Ces protéines présentent une forte similarité de structure malgré une très faible similarité de séquence. Les protéines SsbG s'assemblent en trimères (143) tandis que les protéines Pur α et YP_292156.1 s'assemblent en dimères (142). Les protéines Whirly de plantes accomplissent de nombreuses fonctions liées au métabolisme de l'ADN alors que les protéines Whirly de trypanosomatides sont impliquées dans le phénomène d'édition de l'ARN dans les mitochondries. Les protéines Pur α sont quant à elles impliquées dans la régulation de la transcription et le transport de mRNP au niveau des neurones (142,144). Les protéines SsbG, YP_292156.1 et YP_400729.1 ne lient pas les acides nucléiques (143).

Avant notre étude, seule la structure des protéines MRP1/2 avait été obtenue en complexe avec des acides nucléiques (141). L'élucidation de la structure de MRP1/2 lié à un ARN guide a révélé que MRP1/2 peut lier l'ARN mono- ou bicaténaire avec peu de spécificité de séquence. La liaison à l'ARN se fait au niveau d'un sillon fortement électropositif formé par les feuilletts β de la protéine. Ce sillon comprend de nombreux acides aminés basiques impliqués dans la liaison des phosphates de l'ARN. Ce mode de liaison expose les bases azotées au solvant permettant ainsi la formation d'un duplex d'ARN.

1.2.2 Les stratégies de reconnaissance des acides nucléiques monocaténares

Il existe plusieurs mécanismes différents de liaison des acides nucléiques monocaténares. Ceci est dû au fait que, contrairement aux acides nucléiques bicaténares qui ont une structure relativement rigide, les acides nucléiques monocaténares possèdent une très grande flexibilité étant donné que sept angles de torsion définissent la conformation de chaque nucléotide de la chaîne. Nous allons à présent voir quelques généralités concernant la reconnaissance des acides nucléiques monocaténares.

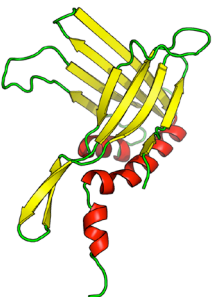
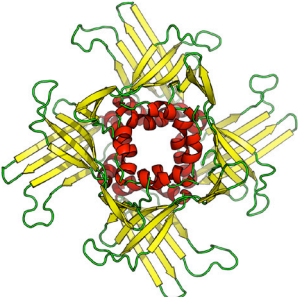
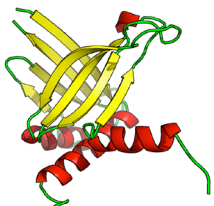
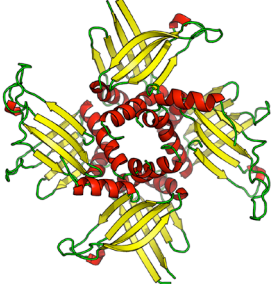
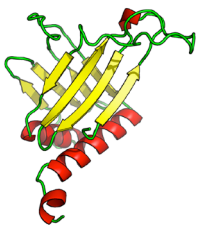

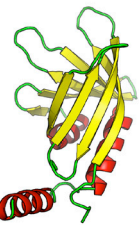

1.2.2.1 Types d'interactions impliquées dans la liaison des acides nucléiques monocaténares

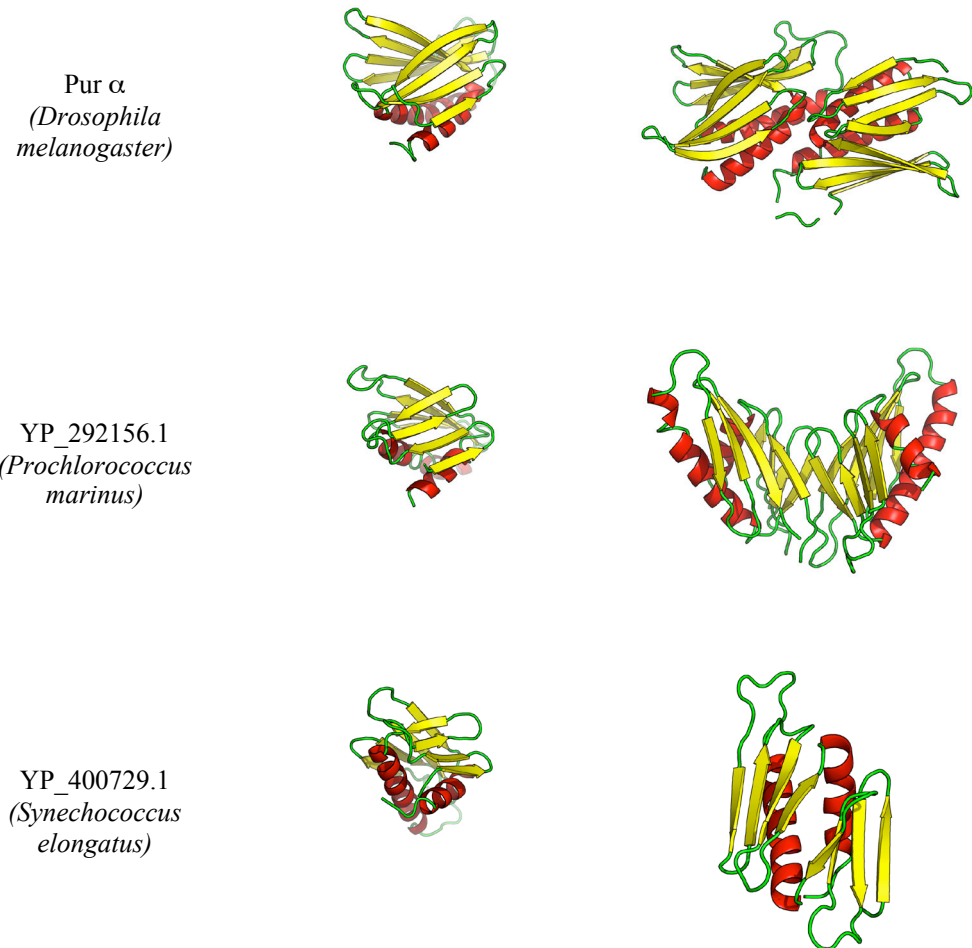
Des interactions électrostatiques s'établissent lorsque des particules de charges contraires sont à proximité. Chez les protéines liant les acides nucléiques monocaténares, on observe fréquemment des interactions électrostatiques entre les résidus basiques des protéines (lysine, arginine et histidine) et les phosphates des acides nucléiques. Ces interactions stabilisent généralement les acides nucléiques de manière séquence-indépendante et augmentent l'affinité de la liaison des protéines aux acides nucléiques (120).

Des empilements aromatique peuvent s'établir entre les acides aminés aromatiques des protéines (phénylalanine, tryptophane, tyrosine et histidine) et les bases azotées des acides nucléiques. On distingue deux types d'empilements aromatique : le contact face-à-face et le contact côté-à-face. Dans le premier cas, les plans des cycles sont parallèles, avec une distance moyenne de 3.3 Å entre les plans, de manière à assurer un recouvrement des nuages aromatiques. Dans le deuxième cas, les groupements aromatiques sont situés à la perpendiculaire l'un de l'autre, de manière à ce qu'il y ait attraction entre le cœur riche en électrons d'un des groupements et la périphérie pauvre en électrons de l'autre groupement (145). Les empilements aromatiques contribuent peu à la spécificité de liaison des acides nucléiques et il y a peu de corrélation entre la présence de ces interactions et la liaison séquence-spécifique des acides nucléiques monocaténares (145).

Les liaisons de van der Waals se font généralement entre les acides aminés hydrophobes (leucine, valine, isoleucine, méthionine) ou comportant une chaîne aliphatique

Tableau 1-IV. Structure tridimensionnelle des protéines à domaine Whirly.

Protéine (Organisme)	Organisation du domaine	Organisation de la protéine
WHY1 (<i>Solanum tuberosum</i>)		
MRP2 (<i>Trypanosoma brucei</i>)		
MRP1 (<i>Trypanosoma brucei</i>)		
SsgB (<i>Thermobifida fusca</i>)		



Les protéines ont été alignées avec DaliLite puis représentées avec Pymol. Les modèles suivants ont été utilisés pour obtenir ces images. St-WHY1 (code pdb 1L3A); Tb-MRP1/2 (code pdb 2GIA); Tf-SsgB (code pdb 3CM1); Dm-Pur α (code pdb 3K44); YP_292156.1 (code pdb 2IT9); YP_400729.1 (code pdb 2NVN). Les hélices α sont colorées en rouge, les feuillet β sont colorés en jaune et les boucles sont colorées en vert.

(lysine, arginine, glutamine) et les groupements hydrophobes des acides nucléiques comme les bases azotées ou les sucres. Ces liaisons contribuent à la spécificité de reconnaissance des acides nucléiques lorsque l'interaction avec une base azotée s'effectue au niveau d'une poche hydrophobe de la protéine (120,145).

Les ponts hydrogènes sont des interactions impliquant le partage d'un atome d'hydrogène entre un donneur et un accepteur. Leur force dépend de la géométrie de liaison, de la nature des groupements accepteurs et donneurs ainsi que de l'environnement dans lequel se situent les donneurs et les accepteurs (146). Isolés, ils contribuent peu à la liaison des protéines aux acides nucléiques. En revanche, les réseaux de ponts hydrogènes peuvent permettre une liaison forte et séquence-spécifique des protéines aux acides nucléiques (145-147).

1.2.2.2 La spécificité de reconnaissance des acides nucléiques monocaténares

Les protéines liant les acides nucléiques monocaténares accomplissent différentes fonctions au sein des cellules. Pour ce faire, certaines protéines lient indistinctement toutes les séquences d'acides nucléiques monocaténares tandis que d'autres reconnaissent une séquence donnée de manière quasi exclusive. Ainsi les protéines liant les télomères reconnaissent habituellement l'ADN monocaténaire avec un très haut degré de spécificité tandis que les protéines impliquées dans la réplication, la recombinaison ou la réparation de l'ADN lient plutôt l'ADN avec une faible spécificité. En général, les protéines qui lient les acides nucléiques monocaténares de manière séquence-indépendante vont interagir avec les phosphates et les sucres des acides nucléiques tandis que les bases azotées seront exposées au solvant. Les protéines qui lient les acides nucléiques monocaténares de manière spécifique vont quant à elles interagir avec les bases azotées. Celles-ci seront soit reconnues par des réseaux de ponts hydrogènes ou liées dans des poches de la protéine, de manière à maximiser la complémentarité de surface et les interactions de van der Waals. Chez les protéines liant les acides nucléiques de manière séquence spécifique, la grande flexibilité des acides nucléiques monocaténares complique le processus de reconnaissance. En effet, si on remplace une base azotée par une autre au sein d'une séquence spécifiquement reconnue, la flexibilité des acides nucléiques monocaténares entraînera soit le remodelage

de l'acide nucléique monocaténaire de manière à établir des interactions compensatoires, soit l'expulsion du nucléotide excédentaire hors du site de liaison (148). Les protéines ont développé différentes stratégies pour lier les acides nucléiques monocaténaires de manière très spécifique ou au contraire de manière peu spécifique. En voici un aperçu :

Les protéines à domaine Puf, comme la protéine Pumilio, consistent en une répétition d'hélices α formant une spirale (120). La liaison de l'ARN se fait sur une face de la spirale. Les bases azotées de l'ARN pointent vers la spirale tandis que les groupements phosphates sont exposés au solvant. Sur la spirale, chaque hélice α reconnaît une seule base d'ARN grâce à un réseau de ponts hydrogènes dirigés vers les groupements fonctionnels des bases azotées. Cette stratégie permet la liaison spécifique de l'ARN. De plus, en mutant les acides aminés situés sur une seule hélice α , il est possible de modifier la spécificité de reconnaissance pour une seule base d'ARN.

Pot1 est une protéine à domaine OB-fold impliquée dans la liaison de l'ADN monocaténaire télomérique chez la levure *Schizosaccharomyces pombe* (149). La liaison de Pot1 à l'ADN télomérique est si spécifique que la mutation d'une seule base dans la séquence reconnue par Pot1 entraîne une diminution d'affinité de plus de 10 000 fois (149). La structure de Pot1 en complexe avec de l'ADN monocaténaire télomérique a révélé que Pot1 maintient l'ADN dans une conformation partiellement repliée. Cette conformation est en partie tributaire d'interactions protéines-ADN mais aussi d'interactions ADN-ADN. Dans ce cas, l'identité des bases azotées est décodée à la fois par des interactions avec la protéine mais aussi avec d'autres bases azotées.

RPA (*Replication Protein A*) est une protéine trimérique impliquée dans la réplication, recombinaison et réparation de l'ADN chez les cellules eucaryotes. Elle contient six domaines OB-fold répartis sur trois chaînes peptidiques : RPA70, RPA32 et RPA14. La liaison à l'ADN se fait de manière séquence indépendante par interaction progressive avec quatre des domaines OB-fold de RPA (122). La structure cristalline de RPA70 en complexe avec un ADN monocaténaire de huit nucléotides révèle que l'ADN interagit avec RPA70 au niveau d'une crevasse formée par les planchers β des deux domaines OB-fold agencés en

tandem (150). Des acides aminés aromatiques et basiques interagissent respectivement avec les bases azotées et les groupements phosphates.

1.2.2.3 La coopérativité dans la liaison des acides nucléiques monocaténaire

La liaison d'une protéine à une longue molécule d'acide nucléique monocaténaire peut influencer la manière dont les prochaines protéines se fixeront à cette même molécule. Si la liaison de la première protéine favorise la liaison des protéines suivantes, on parle alors de coopérativité positive. Dans le cas contraire, on parle de coopérativité négative. Les protéines liant les acides nucléiques monocaténaires possèdent différents degrés de coopérativité et des exemples de coopérativité positive et négative existent (120).

Un exemple emblématique d'une protéine présentant de la coopérativité positive dans sa liaison à l'ADN monocaténaire est la protéine SSB (*Single-Stranded DNA Binding Protein*) d'*Escherichia coli*. Cette protéine à domaine OB-fold forme un homotétramère qui peut lier l'ADN selon différents modes définis par la stœchiométrie de liaison de la protéine SSB à l'ADN monocaténaire (Figure 1-7). Un changement dans la concentration de la protéine ou dans la force ionique du milieu peut entraîner un changement de mode ainsi qu'un changement dans la coopérativité de liaison à l'ADN (120). Dans le mode (SSB)₃₅, chaque tétramère de SSB se lie à 35 nucléotides d'ADN avec une forte coopérativité positive étant donné qu'une protéine SSB a une probabilité 100 000 fois plus forte de se lier à proximité d'une protéine déjà liée à l'ADN que de se lier sur un site isolé (151). Dans le mode (SSB)₃₅, seulement deux sous-unités du tétramère de SSB lient l'ADN monocaténaire. Les deux sous-unités restantes interagissent avec des tétramères de SSB à proximité de manière à former un filament nucléoprotéique (152). La protéine SSB possède également un mode de liaison peu coopératif appelé (SSB)₆₅. Dans ce cas, chaque tétramère de SSB se lie à 65 nucléotides et il n'y a que peu d'interaction entre des tétramères adjacents. Ces deux modes de liaison sont différenciables par microscopie électronique. En effet, lorsque l'ADN est lié par des protéines SSB en mode (SSB)₃₅, le complexe nucléoprotéique a l'apparence d'un filament homogène alors qu'en mode (SSB)₆₅, il a l'apparence d'un collier de perles. On ignore toutefois quel est le rôle de ces deux modes de liaisons *in vivo*.

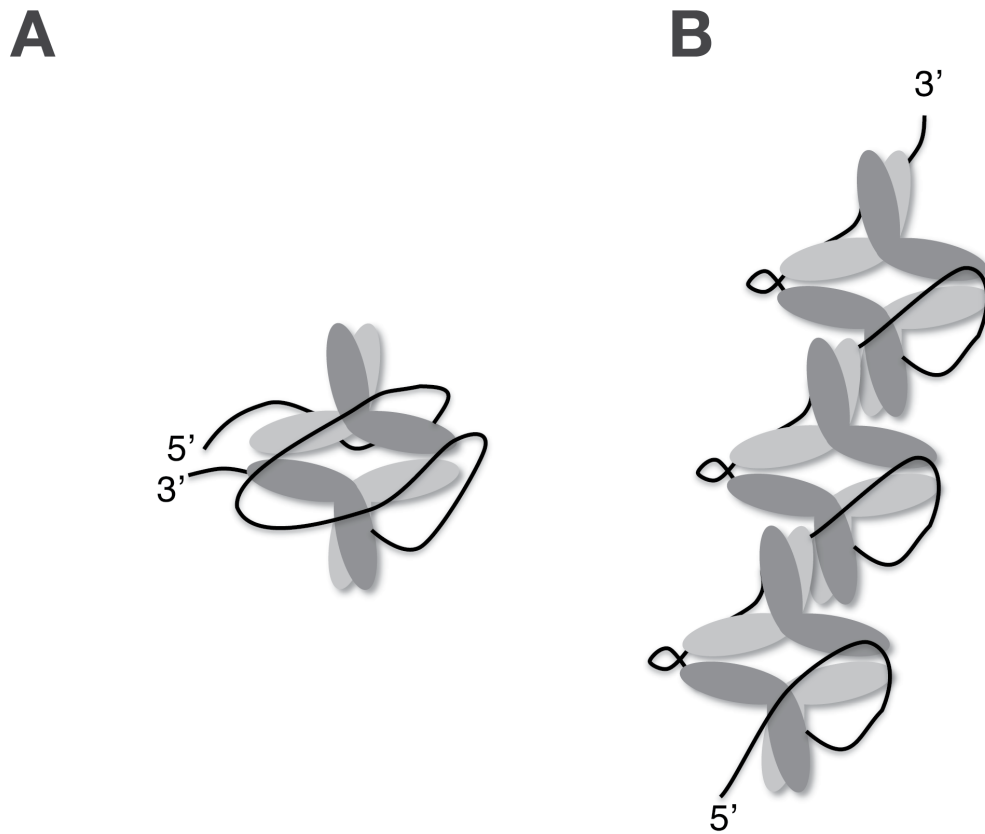


Figure 1-7. Modes de liaison à l'ADN monocaténaire de la protéine SSB d'*Escherichia coli*. (A) Dans son mode $(SSB)_{65}$, la liaison à l'ADN monocaténaire est peu coopérative, les extrémités d'ADN sont situées à proximité l'une de l'autre et les quatre sous-unités de la protéine sont impliquées dans la liaison de l'ADN. (B) Dans son mode $(SSB)_{35}$, la liaison à l'ADN monocaténaire est fortement coopérative, les extrémités d'ADN ne sont pas situées à proximité l'une de l'autre et seulement deux des sous-unités de la protéine sont impliquées dans la liaison à l'ADN. Cette figure est librement adaptée de (120).

Chez le protozaire *Sterkiella nova*, les protéines TEBP- α et TEBP- β sont impliquées dans la liaison des télomères. Ces deux protéines lient l'ADN de manière coopérative en formant un complexe ternaire α - β -ADN. Bien que TEBP- β ne lie que faiblement l'ADN, sa liaison à un complexe α -ADN entraîne une restructuration résultant en une stabilisation du complexe α - β -ADN (153). On parle alors de coopérativité hétérotypique positive des protéines TEBP- α et TEBP- β (120). En revanche, la liaison d'une protéine TEBP- α à une séquence d'ADN comportant plusieurs répétitions d'une séquence télomérique affaiblit la liaison des protéines TEBP- α se fixant aux autres sites (154). On parle alors de coopérativité homotypique négative des protéines TEBP- α (120).

1.2.2.4 La protection ou l'exposition des bases azotées

Certaines protéines vont protéger les bases azotées des acides nucléiques monocaténaire de manière à empêcher la formation de duplex tandis que d'autres protéines vont au contraire les exposer de manière à favoriser leur appariement avec des acides nucléiques de séquences complémentaires. Ainsi, les protéines MRP1/2 permettent l'appariement entre des ARN_{guide} et leur ARN_m correspondant (141). L'ARN bicaténaire ainsi formé sera par la suite pris en charge par un complexe enzymatique, l'éditosome, qui modifiera l' ARN_m en fonction de la séquence de l' ARN_{guide} . Pour favoriser la formation d'un duplex d'ARN à partir de molécules monocaténaire, les protéines MRP1/2 vont principalement lier le squelette de phosphates de l' ARN_{guide} de manière ce que les bases azotées soient exposées au solvant. De plus, l' ARN_{guide} sera stabilisé dans une conformation qui va faciliter son appariement. Finalement, grâce à de nombreux acides aminés chargés positivement, les protéines MRP1/2 vont pouvoir réduire la répulsion électrostatique entre les deux molécules d'ARN complémentaires.

La recombinaise RecA, qui catalyse l'échange de brin entre un ADN bicaténaire « donneur » et un ADN monocaténaire « accepteur », constitue un autre exemple de protéine favorisant l'appariement entre des acides nucléiques. Une étude structurale (155) démontre en effet que RecA forme un filament protéique lors de sa liaison à de l'ADN monocaténaire. Cet ADN est principalement lié au niveau du squelette phosphate, ce qui permet d'exposer les bases azotées. En fait, l'ADN est lié par triplets de nucléotides et il est intéressant de

constater que ceux-ci ont une conformation proche de celle de l'ADN B. Le rôle de la protéine est ainsi de déstabiliser le duplex d'ADN « donneur » de manière à en extraire un triplet d'ADN qui pourra interagir à travers de interactions de type Watson-Crick avec l'ADN monocaténaire « accepteur ». Si le donneur est complémentaire à l'accepteur, un nouveau duplex d'ADN sera formé et la protéine pourra continuer avec le triplet suivant.

1.2.2.5 Dynamisme des protéines liant l'ADN monocaténaire

Les protéines SSB bactériennes sont impliquées dans de nombreux aspects du métabolisme cellulaire et notamment dans les processus de réplication, recombinaison et réparation de l'ADN (156). En plus de leur domaine OB-fold, ces protéines possèdent à leur extrémité C-terminale un motif acide/aromatique séparé du domaine OB-fold par une région désordonnée. Les protéines SSB bactériennes ont longtemps été considérées comme des entités statiques dont le rôle était limité au recouvrement et à la protection de l'ADN. Plusieurs études sont venues changer cette perception. Tout d'abord, il a été montré que le motif acide/aromatique peut se fixer sur une surface électropositive des protéines impliquées dans le métabolisme de l'ADN (157,158), permettant ainsi leur recrutement au niveau de l'ADN. Une étude par transfert d'énergie de fluorescence entre molécules individuelles a également montré que les protéines SSB peuvent glisser sur l'ADN monocaténaire (159). Ce mouvement, qui ne nécessite pas le décrochage complet de la protéine SSB de l'ADN monocaténaire, est permis grâce à la proximité des extrémités d'ADN lors de la liaison des protéines SSB (Figure 1-8). En effet, si un domaine se libère d'une extrémité d'ADN, il lui est alors possible de se lier à l'autre extrémité, occasionnant ainsi un déplacement. Ce déplacement, dont la direction est normalement aléatoire, peut être rendu directionnel si une protéine polymérise sur l'ADN monocaténaire en amont ou en aval de la protéine SSB (159).

1.2.3 Les protéines liant l'ADN monocaténaire dans les organites de plantes

Des protéines liant l'ADN monocaténaire ont été isolées dans les organites de plantes. Vermel et al. (2002) en a détecté plusieurs par chromatographie d'affinité à partir des mitochondries de pomme de terre. Par la suite, deux protéines présentant de l'homologie avec les protéines SSB bactériennes ont été découvertes chez *Arabidopsis* (160) :

At4g11060 (mtSSB1) et At3g18580 (mtSSB2). Ces deux protéines possèdent un peptide signal de localisation vers la mitochondrie et des expériences de fusion GFP ont confirmé que mtSSB1 est principalement localisée au niveau des mitochondries. Cette protéine, qui possède une affinité de liaison à l'ADN monocaténaire comparable à celle des protéines SSB bactériennes, est également capable de stimuler l'activité d'échange de brins de la recombinaison RecA (160) et est importante pour la viabilité des plantes (résultats non publiés dans (46)).

Les protéines OSB constituent une deuxième classe de protéines liant l'ADN monocaténaire chez les organites de plantes. Certaines de ces protéines sont dirigées vers les mitochondries tandis que d'autres sont dirigées vers les chloroplastes (46). Les protéines OSB lient l'ADN monocaténaire sans spécificité de séquence apparente. Elles sont constituées d'un domaine OB-fold semblable à celui des protéines SSB bactériennes et d'un nouveau type de domaine dénommé PDF. C'est ce dernier qui est principalement porteur de l'activité de liaison à l'ADN tandis que le domaine OB-fold permettrait la formation d'oligomères. Des plantes n'exprimant pas la protéine OSB1 développent de la variégation au fil des générations. Celle-ci ne se transmet que maternellement, ce qui suggère la présence d'un trait lié aux organites. Plus précisément, il s'agit de changements dans la stœchiométrie des réarrangements du génome mitochondrial (46).

1.3 Les protéines Whirly de plantes

1.3.1 Les protéines Whirly accomplissent de nombreux rôles reliés au métabolisme des acides nucléiques.

1.3.1.1 Régulation transcriptionnelle du gène *PR-10a*

Les protéines Whirly de plantes ont été découvertes chez *Solanum tuberosum* suite à l'étude de la régulation de l'expression du gène de défense *PR-10a* (*Pathogenesis Related 10a*). Chez la pomme de terre, le gène *PR-10a* est induit suite à la blessure de la plante ou après traitement avec un agent pathogène tel que *Phytophthora infestans* ou avec un éliciteur de la réponse de défense comme l'acide arachidonique. Des expériences de gène

rapporteur ont montré que deux éléments d'ADN situés dans le promoteur du gène *PR-10a* en régulent l'expression (161). La suppression de l'élément ERE (*Elicitor Response Element*) entraîne une diminution de l'expression de *PR-10a* tandis que la délétion de l'élément SE (*Silencing Element*) entraîne une augmentation de l'expression. Des études de retard sur gel ont montré que le ERE est lié par la protéine WHY1 (161,162) tandis que le SE est lié par la protéine SEBF (*Silencing Element Binding Factor* (163)). La protéine WHY1 a été isolée par chromatographie d'affinité à l'élément ERE (162). Une fois la protéine purifiée, sa séquence a été obtenue par spectrométrie de masse et la protéine a pu être clonée. La protéine WHY1 lie le ERE sous sa forme monocaténaire (162) et l'altération de la séquence du ERE provoque une diminution de la liaison avec la protéine WHY1 dans des expériences de retard sur gel (162) ainsi que dans des expériences d'essai rapporteur chez la levure (164). La liaison de WHY1 au ERE a été démontrée *in vivo* chez la pomme de terre grâce à la technique d'immunoprécipitation de la chromatine (164). Cette expérience a montré que WHY1 n'est pas constitutivement liée au ERE mais qu'elle s'y lie après blessure de la plante ou traitement de la plante avec un éliciteur de la réponse de défense.

1.3.1.2 Régulation transcriptionnelle du gène *KPI*

KP1 appartient à la famille des kinésines, une classe de protéines impliquées notamment dans le transport des vésicules et des organites sur les microtubules (165,166). KP1 est fermement liée aux mitochondries ce qui suggère qu'elle pourrait participer au transport des mitochondries sur les microtubules (165). Une étude par gène rapporteur du promoteur du gène *KPI* a révélé qu'une séquence de 43 paires de bases riche en A/T, le KPPE (*KPI-Related Element*), module l'expression de *KPI* (167). Les protéines liant le KPPE ont pu être isolées par chromatographie d'affinité et identifiées par spectrométrie de masse. Il s'agit des protéines WHY1 et WHY3 d'*Arabidopsis* (168). Des expériences d'immunoprécipitation de la chromatine ont montré que WHY1 et WHY3 peuvent lier le KPPE *in vivo* (168). Ces études ont cependant été effectuées avec des plantes qui surexpriment une forme tronquée des protéines WHY1 et WHY3. Cette forme tronquée ne comprend pas le signal de localisation vers les chloroplastes qui est originellement présent chez WHY1 et WHY3.

1.3.1.3 Régulation de la longueur des télomères

Une étude protéomique a permis d'isoler des protéines se liant à des séquences télomériques (TTTAGGG)_n chez *Arabidopsis* (169). Une des protéines identifiées dans cette étude est WHY1. Une étude subséquente s'est intéressée au rôle de cette protéine dans le métabolisme des télomères (170). Dans cette étude, il a été montré que WHY1 lie spécifiquement la séquence télomérique (TTTAGGG)_n. De plus, l'ajout de la protéine WHY1 à une réaction d'extension de télomères entraîne une augmentation de l'activité télomérase. Des plantes dépourvues de la protéine WHY1 possèdent une plus forte activité télomérase que des plantes de type sauvage, ainsi que des télomères plus longs. Au contraire, des plantes qui surexpriment WHY1 possèdent une plus faible activité télomérase que des plantes de type sauvage, ainsi que des télomères plus courts. Les auteurs de cette étude suggèrent que WHY1 régule la longueur des télomères en se liant à l'ADN télomériques et en empêchant sa liaison par les télomérases.

1.3.1.4 Épissage des ARN

Suite à l'immunoprécipitation des protéines interagissant avec CRS1 (*Chloroplast RNA Splicing 1*), une protéine de maïs impliquée dans l'épissage de l'intron de *atpF* dans les chloroplastes, la protéine WHY1 a été isolée et identifiée par spectrométrie de masse (171). La liaison entre WHY1 et CRS1 semble cependant être médiée par de l'ARN. Des expériences d'immunoprécipitation de la chromatine suivies de l'hybridation sur une puce correspondant au génome chloroplastique ont montré que WHY1 lie une sous-population d'ARN ainsi que tout l'ADN chloroplastique *in vivo*. WHY1 contribue à une biogénèse adéquate des chloroplastes étant donné que des plantes de maïs dont le niveau de WHY1 a été diminué grâce à l'insertion d'un transposon dans le gène de *WHY1* sont plus pâles (ivoire ou vert pâle) et possèdent nettement moins de ribosomes chloroplastiques que des plantes de type sauvage. On ne sait toutefois pas comment le manque de Whirly entraîne une diminution de la quantité de ribosomes chloroplastiques.

Une étude récente par le groupe du Dr. Krupinska (172) vient confirmer l'implication des protéines Whirly dans l'épissage de l'ARN chez l'orge, une autre monocotylée. Cette étude révèle que la protéine Hv-WHY1 (*Hordeum vulgare* Whirly 1), un homologue de Zm-

WHY1 (*Zea mays* Whirly 1), se lie aux ARN_m comportant des introns. De plus, chez des plantes d'orge dont l'expression de WHY1 a été réduite, on remarque une diminution dans l'épissage de ces ARN_m. Cette étude diverge cependant de celle du Dr. Barkan sur plusieurs aspects. En effet, aucun phénotype visible n'est observé suite à la réduction du niveau de WHY1 dans les plantes d'orge. Également, des études de sédimentation sur gradient de sucrose ainsi que des études de microscopie indiquent que seulement une petite fraction des protéines WHY1 se lierait à l'ADN au sein des nucleoïdes. Le reste de la protéine serait dans le stroma et ne serait pas liée aux acides nucléiques.

1.3.1.5 Maintien de la stabilité du génome chloroplastique

Pour étudier le rôle des protéines WHY1 et WHY3 dans les organites, des plantes d'*Arabidopsis thaliana* déficientes dans l'expression de ces protéines ont été obtenues (173). Les plantes mutantes pour WHY1 ou WHY3 ne présentent pas de phénotype. Toutefois, lorsque les deux mutations sont combinées dans la même lignée, une petite fraction des plantes (4,6%) présente un phénotype de variégation. La transmission maternelle du phénotype de variégation lors de croisements suggère qu'un défaut dans l'ADN chloroplastique pourrait être à l'origine de la formation des chloroplastes non fonctionnels dans les plantes variégées. De tels défauts ont en effet été mis en évidence et semblent provenir de recombinaisons illégitimes entre de très courtes séquences répétées d'ADN. Ces recombinaisons surviennent aussi, mais à de plus faibles fréquences, chez les plantes de type sauvage. Elle surviennent plus fréquemment chez les plantes ne possédant pas WHY1 ou WHY3, et sont très fréquentes chez les plantes ne possédant aucun des deux gènes. Ces réarrangements sont également présents dans l'ADN de plants de maïs déficients dans l'expression de WHY1.

1.3.2 Phylogénie des protéines Whirly

Les protéines de la famille Whirly sont principalement présentes chez les plantes, mais on les retrouve également chez l'algue verte *Chlamydomonas reinhardtii*, la diatomée *Thalassiosira pseudonana* et le protozoaire *Plasmodium falciparum* (sous le nom de PFE1025c). Des protéines homologues au Whirly sont également présentes

chez des protéobactéries beta et gamma. Il est intéressant de remarquer que les protéines Whirly se retrouvent principalement chez des espèces photosynthétiques. La présence des protéines Whirly chez l'organisme non photosynthétique *Plasmodium falciparum* pourrait alors s'expliquer par le fait que ce protozaire possède un apicoplaste, une organite non-photosynthétique apparentée au chloroplaste des plantes. Cependant, ce ne sont pas toutes les espèces photosynthétiques qui possèdent des protéines Whirly. Ainsi, aucun Whirly n'a été découvert chez la mousse *Physcomitrella patens* ni chez les cyanobactéries. Par contre, des homologues lointains des Whirly sont présent chez les ciliés *Tetrahymena thermophila* et *Paramecium tetraurelia*. Il est également intéressant de remarquer que seuls les Whirly de plantes possèdent un motif contenant des résidus acides/aromatiques à leur extrémité C-terminale. De plus, au sein des protéines Whirly, le domaine Whirly n'est jamais utilisé en conjonction avec un autre domaine.

Certaines espèces possèdent plus d'un gène codant pour une protéine de type Whirly. C'est le cas notamment chez les plantes où au moins deux gènes sont présents. Le produit d'un des gènes est alors dirigé vers la mitochondrie tandis que l'autre est dirigé vers le chloroplaste (174). La présence de plusieurs protéines Whirly dans les organites de certaines espèces, ainsi que la conservation de la région de tétramérisation de la protéine, suggèrent que des hétérotétramères de Whirly pourraient exister (140). On ne connaît toutefois pas de rôles biologiques associés à la formation d'hétérotétramères.

1.3.2.1 TIF1 : un homologue lointain des Whirly

Tetrahymena thermophila est un protozoaire cilié utilisé comme organisme modèle dans l'étude de la réplication de l'ADN chez les eucaryotes. Le génome de cet organisme existe sous deux formes différentes, la forme somatique et la forme germinale, contenues dans deux noyaux différents qui cohabitent dans la même cellule. Un de ces noyaux, le macronoyau, est le site de la transcription des gènes tandis que le deuxième noyau, le micronoyau, permet la propagation du matériel génétique. L'ADN contenu dans le macronoyau provient d'un réarrangement de l'ADN contenu dans le micronoyau. Le génome du micronoyau est diploïde et réparti sur cinq chromosomes. Le génome du macronoyau est quant à lui fortement polyploïde (environ 45 copies) et réparti sur environ deux-cent

cinquante chromosomes (pour revue, se référer à (175)). La division du génome en deux noyaux se fait lors de la reproduction sexuelle par conjugaison de deux cellules. Il y a alors destruction des macronoyaux parentaux, formation d'un nouveau micronoyau à partir des micronoyaux parentaux et création d'un nouveau macronoyau à partir du micronoyau.

Toujours chez *Tetrahymena thermophila*, de nombreuses études ont été faites sur la réplication de son minichromosome codant pour l'ARN 35S ribosomique. Ce minichromosome est amplifié plus de 10 000 fois durant la formation du macronoyau. Des études ont montré que deux types de séquences d'ADN contribuent particulièrement à la réplication du minichromosome. Ces séquences ont été nommées élément de type I et élément de pause. La mutation de ces éléments provoque des anomalies dans le processus de réplication du minichromosome. Des études par retard sur gel ont montré que l'élément de type I est lié par les protéines TIF1, TIF2, TIF3 et TIF4. Ces protéines ont la particularité de reconnaître l'ADN sous sa forme monocaténaire. TIF1 est une protéine de 21,5 kDa qui s'agence en tétramère (176) et qui possède 18% d'identité de séquence et 40% de similarité de séquence avec la protéine WHY1 de pomme de terre (177). La région la plus conservée est située à l'extrémité C-terminale de la protéine et correspond aux deux dernières hélices α des protéines Whirly de plantes.

Des individus dont l'expression de TIF1 a été altérée ont été générés de manière à comprendre le rôle physiologique de la protéine TIF1 (178). Ces individus se révèlent plus sensibles aux dommages à l'ADN. Une étude plus détaillée montre que TIF1 est en fait requis pour activer un point de contrôle dépendant du complexe ATR durant la phase S (179). En plus de ce rôle, TIF1 est requis pour activer les origines de réplication au niveau du macronoyau au début de la phase S (178).

1.3.3 Structure des protéines Whirly de plantes et mécanisme de liaison à l'ADN

L'élucidation de la structure cristalline de WHY1 de pomme de terre (140,180) a été un pas en avant vers une meilleure compréhension du mode de liaison à l'ADN monocaténaire des protéines Whirly de plantes. En analysant la liaison à l'ADN de

protéines WHY1 tronquées, l'activité de liaison à l'ADN a pu être localisée au sein du domaine Whirly de la protéine (140). En comparant la structure de la protéine WHY1 avec celle d'autres protéines liant l'ADN monocaténaire, il a été proposé que les feuillets β incurvés constituent la surface de liaison à l'ADN monocaténaire (Figure 1-8). En effet, de nombreux résidus aromatiques ou basiques sont situés sur une des faces des planchers β du domaine Whirly. Ceux-ci pourraient être impliqués dans la liaison des acides nucléiques monocaténaires (140). Un motif KGKAAL situé en extrémité du premier plancher β semble être important pour la liaison à l'ADN étant donné que la mutation de ce motif en QGQGGV provoque une altération de la liaison à l'ADN sans modifier l'agencement tétramérique de la protéine (140). Le motif YDW ainsi que la lysine 188 semblent également importants pour la liaison de WHY1 à l'ADN monocaténaire (résultats non publiés dans (181)). Chez WHY1 d'*Arabidopsis*, deux mutations ponctuelles ont été étudiées à la fois *in vitro* et *in vivo*. La mutation P183S est localisée dans le plancher β potentiellement impliqué dans la liaison de l'ADN monocaténaire tandis que la mutation G148E est localisée dans une hélice α permettant la formation de tétramères. Des études par retard sur gel ont montré que ces mutations diminuent la liaison à l'ADN *in vitro*, tandis qu'*in vivo* des plantes d'*Arabidopsis* exprimant des protéines WHY1 comportant ces mutations sont plus susceptibles à des infections par des pathogènes (164).

Finalement, la structure cristallographique de WHY1 a révélé que le motif acide/aromatique situé à l'extrémité C-terminale de la protéine interagit avec les feuillets β de la protéine à l'emplacement proposé du site de liaison aux acides nucléiques monocaténaires (Figure 1-9 et (140)). La mutation de résidus de l'extrémité C-terminale entraîne une augmentation de l'affinité de WHY1 pour l'ADN monocaténaire (résultats non publiés dans (181)) ce qui suggère que la séquence C-terminale régule la liaison de WHY1 aux acides nucléiques monocaténaires, probablement en entrant en compétition avec l'ADN monocaténaire pour la liaison aux feuillets β .

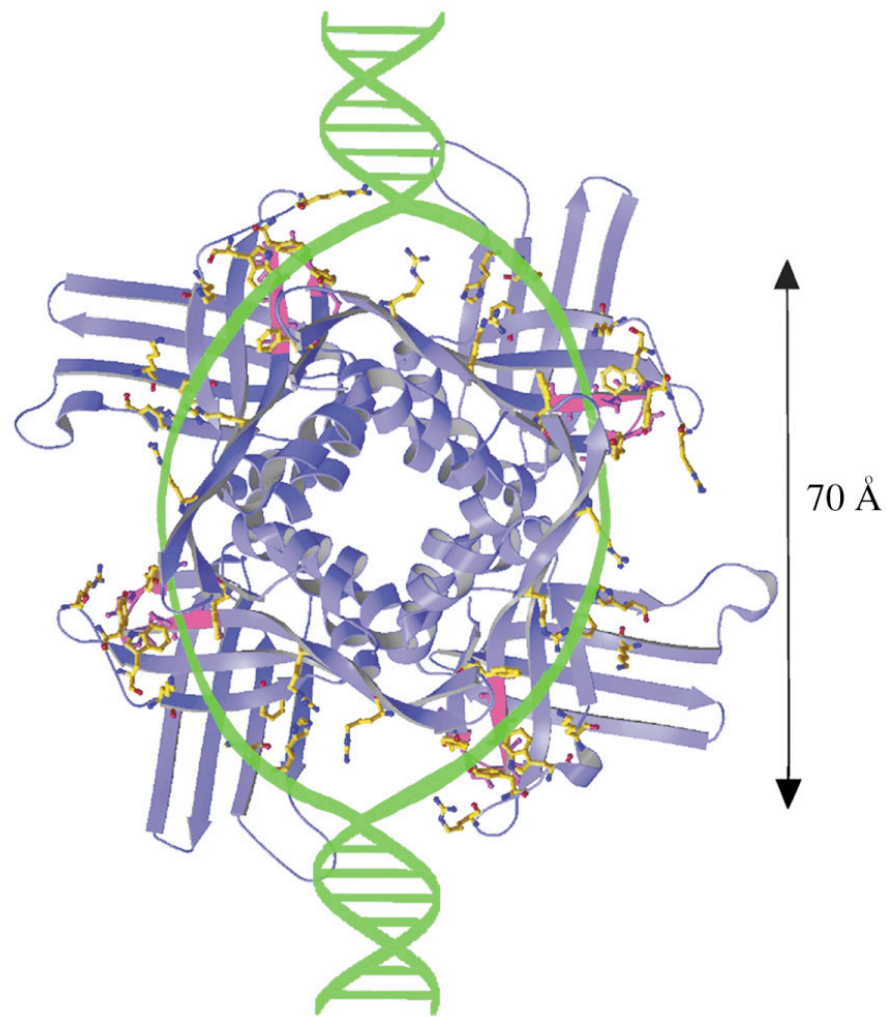


Figure 1-8. Modèle proposé pour la liaison de la protéine WHY1 de pomme de terre à l'ADN monocaténaire suite à l'élucidation de la structure de WHY1 en forme libre (code pdb 1L3A). La protéine est représentée en mauve tandis que l'ADN est représenté en vert. Des résidus basiques ou aromatiques situés au niveau des feuillets β sont représentés en jaune. Cette figure est adaptée de (140) avec la permission de Macmillan Publishers Ltd (copyright 2002).

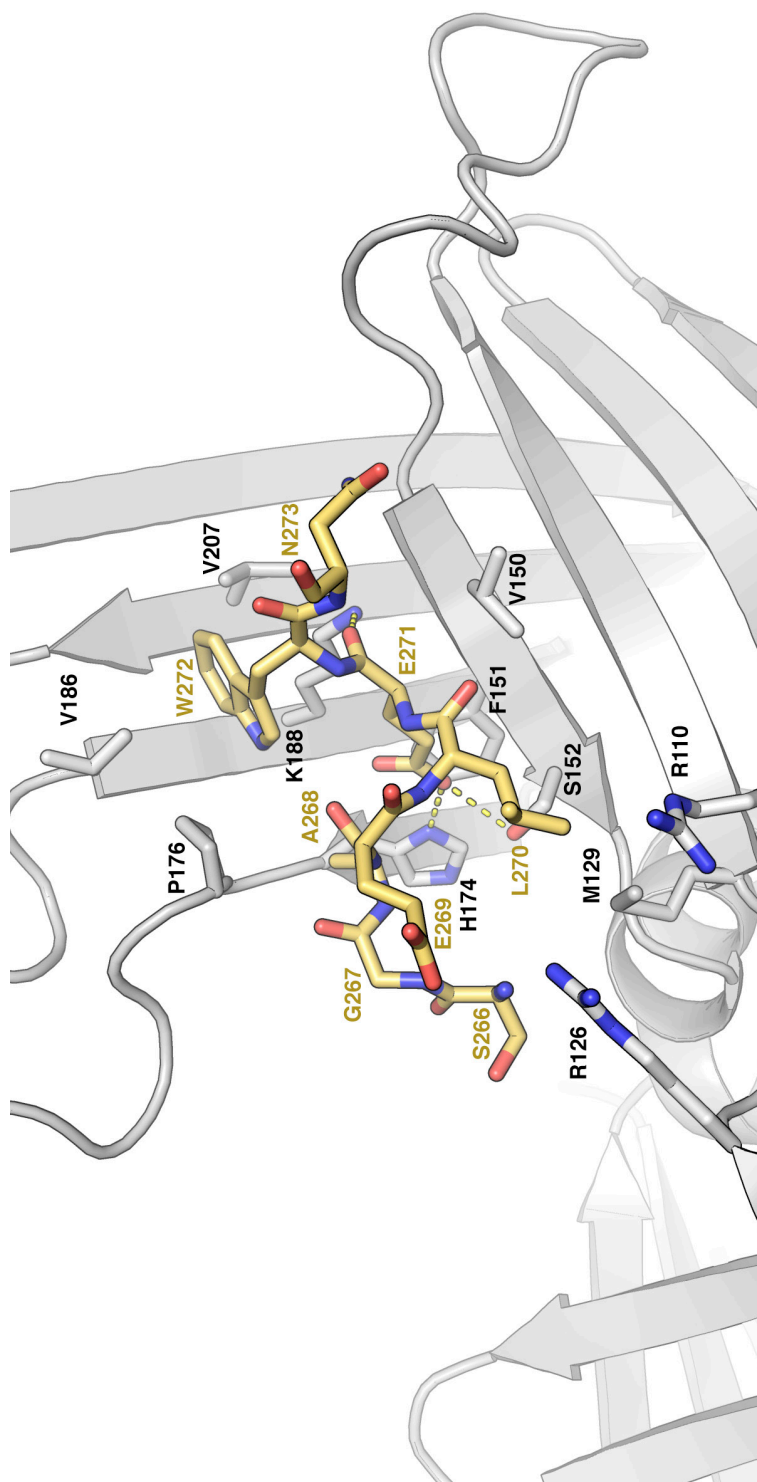


Figure 1-9. Interaction entre le domaine Whirly de WHY1 et le motif acide/aromatique C-terminal. Le domaine Whirly de la protéine WHY1 (code pdb 1L3A) est coloré en gris tandis que les atomes de carbones du motif acide/aromatique C-terminal sont colorés en jaune.

1.4 Objectifs du projet de recherche et justification des méthodes employées

1.4.1 Objectifs du projet de recherche

Chez *Arabidopsis*, nous avons vu qu'en absence des protéines WHY1 et WHY3 il y a production et accumulation de réarrangements de l'ADN qui s'apparentent à ceux qui sont formés lors de la réparation de l'ADN par la voie de la réplication induite par les cassures dépendantes de très courtes séquences répétées (*MMBIR*). Un premier objectif de notre projet de recherche a été de vérifier si les protéines Whirly sont impliquées dans la réparation de l'ADN.

Toutes les études effectuées sur les protéines Whirly de plantes ont montré une interaction entre ces protéines et les acides nucléiques. La très grande majorité des études ont établi qu'elles reconnaissent l'ADN monocaténaire, même si certaines études ont montré qu'elles peuvent également lier l'ARN ou l'ADN bicaténaire. De nombreuses séquences cibles, souvent différentes, ont été proposées sans que l'on comprenne comment les protéines Whirly se lient à ces séquences. Le deuxième objectif de ce projet de recherche vise à comprendre les bases structurales de la liaison des protéines Whirly de plantes aux acides nucléiques et plus précisément, à l'ADN monocaténaire. Nous avons également cherché à savoir si les protéines Whirly peuvent s'auto-assembler de manière à lier l'ADN de manière coopérative et nous avons analysé le lien entre le mécanisme de liaison aux acides nucléiques et les fonctions biologiques des protéines Whirly, en particulier dans le cadre de la réparation de l'ADN.

1.4.2 Justification des méthodes employées

1.4.2.1 Induction de cassures bicaténaires de l'ADN dans les organites

Pour comprendre le rôle des protéines Whirly dans la réparation des cassures bicaténaire de l'ADN nous avons utilisé une technique mise au point dans notre laboratoire par Étienne Lepage et Alexandre Maréchal qui consiste à traiter des plantes d'*Arabidopsis* avec

des inhibiteurs des ADN gyrases. Ces protéines produisent du surenroulement négatif dans l'ADN. Pour ce faire, elles lient une boucle d'ADN et introduisent une cassure bicaténaire à l'endroit où les brins d'ADN se croisent. Elles catalysent ensuite la traversée du segment intact par le segment clivé pour finalement ressouder la cassure (pour revue, se référer à (182)). L'inhibition de ces protéines par les quinolones comme la ciprofloxacine empêche la soudure de l'ADN, entraînant ainsi la formation de cassures bicaténaires (183-186) au niveau des organites de plantes où les ADN gyrases se localisent (187). Les réarrangements de l'ADN, résultant de la réparation de ces cassures, sont détectés par PCR ainsi que par des buvardages de type Southern.

1.4.2.2 Élaboration de la structure tridimensionnelle des protéines Whirly par radiocristallographie

En principe, deux techniques peuvent être employées afin d'élucider le mécanisme de liaison à l'ADN monocaténaire des protéines Whirly de plantes à l'échelle atomique : la résonance magnétique nucléaire et la radiocristallographie. La première technique est limitée par la taille de l'objet d'étude et un complexe Whirly-ADN (environ 90 000 Da) est trop grand pour pouvoir être étudié par cette technique. Nous avons donc décidé de procéder par radiocristallographie.

La radiocristallographie est une technique permettant de déterminer la structure tridimensionnelle d'une molécule d'intérêt avec une résolution quasi-atomique. Cette technique se base sur le fait qu'un cristal possède la capacité de diffracter un faisceau de rayons X et que l'intensité des rayons X diffractés dépend de l'organisation des atomes au sein d'une maille cristalline. Les principes de cette technique ont été élucidés il y a environ un siècle et la première utilisation de cette technique pour déterminer la structure d'une protéine remonte au début des années 60. Nous allons maintenant voir comment élucider la structure d'une protéine d'intérêt par radiocristallographie. Cette section se base sur l'excellent livre *Biomolecular Crystallography* de Bernhard Rupp (188).

La première étape vers l'élucidation de la structure d'une protéine d'intérêt consiste en la cristallisation de cette protéine. Pour ce faire, celle-ci est fréquemment surexprimée dans un système bactérien et purifiée à homogénéité par chromatographie. Très souvent, la cristallisation de la protéine se fait par la technique de la goutte pendante avec diffusion

de vapeur. Ce dispositif est constitué d'un récipient et de son couvercle. Environ 500 μl d'une « solution mère » contenant tous les ingrédients nécessaires à la cristallisation de la protéine sont versés dans le récipient. On dépose ensuite une petite quantité de la solution contenant la protéine d'intérêt (de 1 à 10 μl) sur le couvercle et on y ajoute une quantité égale de solution mère. Le couvercle est ensuite refermé ce qui permet une équilibration entre la solution mère et la « goutte pendante ». Durant l'équilibration, la concentration de protéine et de solutés vont augmenter dans la goutte pendante amenant ainsi la protéine dans un état de supersaturation qui permet la formation de noyaux de cristallisation. Ces noyaux vont par la suite pouvoir croître et donner naissance à des cristaux. La composition de la solution mère est obtenue empiriquement en testant de très nombreuses recettes.

Les cristaux obtenus sont ensuite utilisés dans des expériences de diffraction des rayons X. Il existe deux sources principales de rayons X : les générateurs à anode tournante et les synchrotrons. Dans les deux cas, on retrouve un même dispositif expérimental où le cristal est maintenu par capillarité dans une boucle en nylon. Cette boucle est elle-même fixée à un goniomètre qui permet de modifier la position du cristal par rapport au faisceau de rayons X. Un flux d'azote gazeux pointé vers la boucle permet de maintenir le cristal à une température de 100 kelvins afin de minimiser les dommages oxydatifs dus à l'exposition de l'échantillon aux rayons X. Durant l'expérience, le cristal est exposé aux rayons X et des patrons de diffraction correspondant à différentes orientations du cristal sont collectés. Bien qu'une transformée de Fourier permette de convertir l'intensité et la phase des faisceaux diffractés en une carte qui indique la densité électronique en fonction des coordonnées cartésiennes, cette opération ne peut pas se faire directement étant donné que les phases des faisceaux diffractés ne peuvent pas être mesurées lors des expériences de diffraction. Ceci donne origine au « problème des phases ». Il est possible de contourner ce problème grâce aux techniques de « remplacement isomorphe » ou de « diffraction anormale ». Aussi, si une protéine possède une structure proche de la protéine d'intérêt, il est possible d'utiliser cette structure pour effectuer un « remplacement moléculaire ». Dans ce cas, on va positionner adéquatement cette structure dans la maille cristalline. Une carte de densité électronique peut alors être obtenue par une transformée de Fourier en utilisant les intensités obtenues expérimentalement et les phases calculées à partir du modèle. Cette carte comporte des attributs provenant de la structure d'origine et des attributs provenant du modèle. À l'aide

d'un ordinateur, on procède alternativement à la construction du modèle dans la densité électronique et à l'affinement du modèle. Au cours de l'affinement, les coordonnées des atomes sont modifiées de manière à améliorer la corrélation entre le modèle et les données expérimentales. De nouvelles cartes de densités sont alors calculées à partir du nouveau modèle permettant d'entamer un nouveau cycle de construction-affinement. Le produit final de la modélisation est un « fichier PDB » qui contient les coordonnées cartésiennes des atomes modélisés ainsi qu'un paramètre de température qui mesure la déviation des atomes autour d'une position centrale et un pourcentage d'occupation utilisé lorsqu'un ligand ne se lie pas à toutes les protéines d'un cristal.

Deux indicateurs statistiques, « R » et « R_{free} », permettent d'évaluer la qualité des modèles. En effet, une fois les données de diffraction collectées, un certain pourcentage (généralement de 5 à 10 %) des réflexions sont marquées de manière à ce qu'elles ne soient pas utilisées au cours des affinements. Le facteur « R » mesure la corrélation entre le modèle et les données expérimentales non marquées tandis qu'un indicateur « R_{free} » mesure la corrélation entre le modèle et les données expérimentales marquées. Si le modèle est biaisé, il y aura une différence importante entre le facteur R et le facteur R_{free} . Il existe d'autres indicateurs qui permettent d'évaluer la qualité des modèles. Une fois un modèle obtenu, il sera analysé. On pourra alors utiliser le modèle pour expliquer des données ou pour émettre de nouvelles hypothèses.

1.4.2.3 Élucidation de l'agencement quaternaire des protéines Whirly par radiocristallographie et microscopie à force atomique

Pour étudier l'agencement quaternaire des protéines Whirly, nous pouvons combiner les techniques de radiocristallographie, de tamis moléculaire et de microscopie à force atomique. La radiocristallographie permet de visualiser, à l'échelle quasi-atomique, les agencements quaternaires des protéines. En effet, au cours de sa cristallisation, une protéine d'intérêt est amenée à interagir avec elle-même. Bien que dans de nombreux cas ces interactions autodirigées ne semblent pas jouer de rôle *in vivo*, il peut arriver que la protéine d'intérêt s'agence *in vivo* en homooligomères et que le processus de cristallisation exploite cette capacité de la protéine à interagir avec elle-même. La technique de tamis moléculaire permet quant à elle d'évaluer la taille des molécules d'intérêt. À partir de cette information,

il est possible de déterminer l'état de l'oligomérisation de la protéine d'intérêt. Finalement, la technique de microscopie à force atomique permet d'obtenir une carte topographique d'un échantillon biologique déposé sur un support de mica. À partir de cette carte, il est possible d'obtenir la hauteur de protéines déposées sur le support de mica et ainsi de déterminer l'état oligomérique de la protéine.

CHAPITRE 2

Purification, cristallisation et analyse préliminaire par diffraction aux rayons X du domaine Whirly de St-WHY2 en complexe avec de l'ADN monocaténaire.

Article 1 : Purification, crystallization and preliminary X-ray diffraction analysis of the Whirly domain of St-WHY2 in complex with single-stranded DNA

Reprinted with permission from the *International Union of Crystallography*
(<http://journals.iucr.org/>)

**Purification, crystallization and preliminary X-ray
diffraction analysis of the Whirly domain of St-WHY2 in
complex with single-stranded DNA**

Laurent Cappadocia, Jurgen Sygusch and Normand Brisson

Department of Biochemistry
Université de Montréal
PO Box 6128, Station Centre-Ville
Montréal, Québec
H3C 3J7, Canada

Acta Cryst. (2008). F64, 1056–1059

2.1 Mise en contexte

L'obtention de la structure tridimensionnelle d'une protéine Whirly en complexe avec de l'ADN monocaténaire par la technique de radiocristallographie est tributaire de l'obtention de cristaux diffractant à haute résolution. Dans un premier temps, nous avons obtenu des cristaux d'un complexe St-WHY1-ADN diffractant de manière si anisotrope qu'il ne nous a pas été possible de résoudre la structure du complexe. En revanche, en utilisant St-WHY2, un paralogue de St-WHY1, nous avons été en mesure d'obtenir des données de diffractions permettant l'élucidation de la structure tridimensionnelle d'un complexe Whirly-ADN.

Les résultats de ce chapitre ont été publiés dans la revue *Acta Crystallographica Section F* en novembre 2008. J'ai optimisé la diffraction des cristaux de St-WHY1-ADN, j'ai obtenu les cristaux de St-WHY2 et de St-WHY2-ADN, j'ai collecté les données de diffraction et j'ai rédigé le présent manuscrit qui a par la suite été révisé par tous les auteurs.

2.2 Abstract

St-WHY1 and St-WHY2 are members of the Whirly family of single-stranded DNA (ssDNA) binding proteins. To understand the mode of binding of the Whirly proteins to single-stranded DNA, crystals of the Whirly domains of both St-WHY1 and St-WHY2 in complex with single-stranded DNA were obtained by the hanging-drop vapour-diffusion method. The diffraction patterns of the St-WHY1-ssDNA complex crystals displayed severe anisotropy and were of low resolution, making them unsuitable for structure determination. In contrast, the crystals of the St-WHY2-ssDNA complex diffracted isotropically to 2.20 Å resolution. The crystallization and data collection to 2.20 Å resolution of St-WHY2 in the free form are also reported.

2.3 Introduction

Single-stranded DNA (ssDNA) binding proteins play a pivotal role in the metabolism of DNA. They are involved in replication, repair, recombination and transcription processes

as well as in protection of the telomeres. In accordance with their important role, ssDNA-binding proteins are ubiquitously found in all realms of life.

The Whirlies belong to a family of ssDNA-binding proteins that are found mainly in the plant kingdom (Desveaux *et al.*, 2002b, 2005). They fulfil multiple roles including regulation of transcription (Desveaux *et al.*, 2000, 2004), protection of telomere ends (Yoo *et al.*, 2007) and possibly replication (Maréchal *et al.*, 2008). These roles are dependent on the ssDNA-binding activity of the Whirly proteins, which is located on the Whirly domain, a 160-amino-acid strongly conserved region that is necessary and sufficient for high-affinity DNA binding (Desveaux *et al.*, 2005). Elucidation of the crystal structure of St-WHY1 (*Solanum tuberosum* Whirly 1) in the free form revealed a tetrameric quaternary structure for the Whirly domain resembling a whirligig (Desveaux *et al.*, 2002b). Despite the elucidation of the structure of St-WHY1 in the free form, the mode of ssDNA binding to the tetramer has remained elusive.

Members of the Whirly family of proteins have been reported to bind different ssDNA sequences. St-WHY1 binds the elicitor response element (ERE; Desveaux *et al.*, 2000) in the promoter region of the *PR-10a* defence gene as a prerequisite for transcriptional activation (Desveaux *et al.*, 2004). At-WHY1 (*Arabidopsis thaliana* Whirly 1), a close homologue of St-WHY1, has been reported to bind the ERE sequence (Desveaux *et al.*, 2004) as well as a sequence-unrelated telomeric repeat (Yoo *et al.*, 2007). Finally, At-WHY2 and St-WHY2 have been reported to bind ssDNA with little sequence specificity (Maréchal *et al.*, 2008; Vermel *et al.*, 2002). In order to understand the structural basis for ssDNA binding, we sought to determine the crystal structure of a complex between a Whirly protein and ssDNA. Here, we report the crystallization of St-WHY1 and St-WHY2 in complex with ERE₃₂, a 32-mer oligonucleotide based on the ERE sequence. Crystals of the St-WHY1-ERE₃₂ complex diffracted anisotropically to low resolution, whereas crystals of the St-WHY2-ERE₃₂ complex diffracted isotropically to 2.20 Å resolution, thereby allowing the elucidation of the crystal structure of a Whirly bound to single-stranded DNA. To gain further insight into the ssDNA-binding mechanism, we also obtained crystals of St-WHY2 in the free form which diffracted to 2.20 Å resolution.

2.4 Materials and methods

2.4.1 Cloning, expression and purification

Recombinant St-WHY1 was expressed and purified as described previously (Desveaux *et al.*, 2002a). Briefly, a pET-21a (Novagen) expression plasmid containing both St-WHY1 (68-273) and a C-terminal hexahistidine tag was transformed into *Escherichia coli* strain BL21 (DE3). The cells were grown at 310 K in Luria-Bertani broth. When the cells reached an OD₆₀₀ of 0.6, isopropyl β-D-1-thiogalactopyranoside was added to a final concentration of 1 mM. After further cell growth for 3 h at 310 K, the cells were harvested, resuspended in 20 mM sodium phosphate pH 7.5, 500 mM NaCl, 25 mM imidazole and lysed by alumina grinding. The recombinant protein was purified by applying the supernatant from the cell lysate onto a HiTrap Chelating nickel-affinity column (GE Healthcare).

St-WHY1Δ (87-253) was obtained by PCR using an St-WHY1 construct as a template and using the primers St-WHY1ΔFwd, GGAATTCCATATGGCATCTACGCCTAAGGT, and St-WHY1ΔRvs, CCGCTCGAGAGGCTTGAAGGAATTTAC. In order to clone St-WHY2, total cDNA from *S. tuberosum* was obtained using the RevertAid First Strand cDNA Synthesis Kit (Fermentas). The gene coding for St-WHY2 (48-216) was then amplified by the polymerase chain reaction with *Pwo* polymerase (Roche) and the primers St-WHY2Fwd, GGAATTCCATATGGCTGATGCAGGAAAACGG, and St-WHY2Rvs, CCGCTCGAGAGGCCGATTAGTGAAGCG. The PCR products were subcloned into a pET-21a vector (Novagen) using *NdeI/XhoI* restriction sites. This cloning strategy added a methionine at the N-terminus of the proteins and a hexahistidine tag (LEHHHHHH) at the C-terminus. Both St-WHY2 and St-WHY1Δ were purified as described for St-WHY1.

2.4.2 Crystallization of St-WHY1-ERE₃₂ and data collection

The oligonucleotide ERE₃₂ (5'-TGTCATTTTTGTCATTTTTGTCATTTTTGTCA-3') used for crystallization was obtained from IDT (Integrated DNA Technologies, Coralville, Iowa, USA). St-WHY1 was mixed with ERE₃₂ in a 1:2 tetramer:ERE₃₂ molar ratio. The complex was allowed to form for 30 min at 293 K and was purified using a Superdex 200

16/60 size-exclusion column (GE Healthcare) pre-equilibrated in a buffer containing 25 mM citrate-phosphate buffer pH 5.5 and 300 mM NaCl. The purified complex was then concentrated to 10 mg ml⁻¹ using Millipore 10K concentrators. The protein concentration was determined using the bicinchoninic acid (BCA) protein assay kit (Pierce). Crystals were typically obtained at 296 K by mixing 3 µl purified complex solution with 3 µl reservoir solution. The reservoir solution (500 µl) contained 0.5-1%(v/v) PEG 550, 500-900 mM NaCl and 25 mM sodium citrate/sodium phosphate buffer pH 5. Crystals appeared within a day but reached their maximal size after a week. Crystals of St-WHY1-ssDNA were transferred into a cryoprotectant composed of the mother liquor with 20%(v/v) ethylene glycol. The crystals were then mounted in a cryoloop (Hampton Research) and flash-cooled in a stream of nitrogen gas at 100 K. 360 frames were recorded using an oscillation range of 0.5° and a crystal-to-detector distance of 350 mm. Diffraction data were collected using an ADSC Quantum 315 CCD detector on beamline X29 at the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory (BNL, USA). The data were processed, indexed and scaled using *HKL-2000* (Otwinowski & Minor, 1997).

2.4.3 Crystallization of St-WHY2 in free form and data collection

After nickel-affinity chromatography, St-WHY2 was further purified using a Superdex 200 16/60 size-exclusion column (GE Healthcare) pre-equilibrated in a buffer containing 10 mM Tris-HCl pH 8.0 and 100 mM NaCl. The purified protein was then concentrated to 20 mg ml⁻¹ using Millipore 10K concentrators. The protein concentration was determined using the BCA protein assay kit (Pierce). Crystals were obtained at 296 K by mixing 3 µl St-WHY2 protein solution with 3 µl reservoir solution. The reservoir solution (500 µl) contained 21%(v/v) PEG 1000, 400 mM (NH₄)₂HPO₄ and 100 mM MOPS pH 7.0. Crystals appeared within a day but reached their maximal size after a week. A single crystal of St-WHY2 in the free form was mounted in a cryoloop (Hampton Research) and flash-cooled in a stream of nitrogen gas at 100 K. 360 frames were recorded using an oscillation range of 0.5° and a crystal-to-detector distance of 325 mm. Diffraction data were collected as above but on beamline X25. The data were processed, indexed and scaled using *HKL-2000* (Otwinowski & Minor, 1997).

2.4.4 Crystallization of St-WHY2-ERE₃₂ and data collection

Purified St-WHY2 was mixed with ERE₃₂ in a 1:2 tetramer:ERE₃₂ molar ratio. The complex was allowed to form for 30 min at 293 K and was purified using a Superdex 200 16/60 size-exclusion column (GE Healthcare) pre-equilibrated in a buffer containing 10 mM Tris-HCl pH 8.0 and 100 mM NaCl. The purified complex was then concentrated to 10 mg ml⁻¹ using Millipore 10K concentrators. The protein concentration was determined using the BCA protein assay kit (Pierce). Crystals were obtained at 296 K by mixing 3 µl purified complex solution with 3 µl reservoir solution. The reservoir solution (500 µl) contained 13.5%(v/v) PEG 6000, 1 M LiCl and 100 mM Tris-HCl pH 8.0. Crystals appeared within a week but reached their maximal size after a month. To reduce the salt concentration, a single crystal of St-WHY2-ERE₃₂ was quickly soaked in a solution containing 13.5%(v/v) PEG 6000, 0.5 M LiCl and 100 mM Tris-HCl pH 8.0. The crystal was then mounted in a cryoloop (Hampton Research) and flash-cooled in a nitrogen-gas stream at 100 K. 360 frames were recorded using an oscillation range of 0.5° and a crystal-to-detector distance of 275 mm. Diffraction data were collected as described for St-WHY1-ERE₃₂. The data were processed, indexed and scaled using *HKL-2000* (Otwinowski & Minor, 1997).

2.5 Results and discussion

The St-WHY1-ERE₃₂ complex initially appeared to be a promising candidate for crystallization since the crystal structure of St-WHY1 had been solved in the free form (Desveaux *et al.*, 2002a,b) and the interaction of St-WHY1 with ERE had been characterized both *in vitro* and *in vivo* (Desveaux *et al.*, 2004). We obtained our initial crystallization conditions for the St-WHY1-ERE₃₂ complex by sparse-matrix screening. Refinement of the conditions led us to obtain thin plate-shaped crystals that appeared in less than 1 d (Fig. 1a). A data set could be collected from these crystals but was unusable owing to severe anisotropy, with diffraction limited to 3 Å along the *a** and *b** dimensions and to 6 Å along the *c** dimension (Fig. 1b). Screening different sequences and lengths of oligonucleotides did not reduce the anisotropy (data not shown). Truncated St-WHY1Δ protein lacking the flexible N- and C-terminus tails yet retaining high ssDNA-binding affinity also did not improve the data anisotropy or increase the resolution (data not shown). To overcome the data anisotropy,

we decided to crystallize the Whirly domain of St-WHY2, a homologue of St-WHY1, both in the free form and as a complex with ssDNA.

Initial crystallization trials for St-WHY2 in the free form and for the St-WHY2-ERE₃₂ complex took place using the high-throughput batch crystallization facility at the Hauptman-Woodward Institute in Buffalo (Luft *et al.*, 2003) and resulted in 17 and 20 hits, respectively. Several crystallization conditions were successfully reproduced using hanging drops, yielding bipyramidal crystals for both St-WHY2 in the free form and St-WHY2-ERE₃₂. Importantly, the crystals of the St-WHY2-ERE₃₂ complex (Fig. 2*a*) diffracted isotropically to 2.20 Å resolution (Fig. 2*b*), whereas the same conditions yielded no crystals for the St-WHY1-ERE₃₂ complex. Interestingly, St-WHY2 in the free form and St-WHY2-ERE₃₂ gave morphologically identical crystals that belonged to space group *F*432, even though they were crystallized in different chemical environments. Crystals were harvested, washed three times with their respective mother liquor, dissolved and then analyzed by 15% SDS-PAGE followed by silver staining (Fig. 3). This confirmed that the protein was present in the crystals of St-WHY2 in the free form and that the St-WHY2-ERE₃₂ crystals contained both the protein and the oligonucleotide.

The Whirly domain of St-WHY2 shares 51% sequence identity and 73% sequence similarity to the Whirly domain of St-WHY1. We therefore used the structure of St-WHY1 (PDB code 1l3a) as a molecular-replacement template for St-WHY2 both in the free form and in the ERE₃₂ complex. The molecular replacement was successful and electron density corresponding to DNA bases was observed in an $F_o - F_c$ electron-density map for the St-WHY2-ERE₃₂ complex but not for St-WHY2 in the free form. Structure refinement is ongoing and should provide the structural basis for the binding of the Whirly proteins to ssDNA.

2.6 Acknowledgements

The contribution of Katia Abdeladim is greatly acknowledged. Thanks also to the Hauptman-Woodward Institute for determining the initial crystallization conditions. Research carried out at the National Synchrotron Light Source, Brookhaven National Laboratory was supported by the US Department of Energy, Division of Materials Sciences and Division

of Chemical Sciences. Assistance by the X25 and X29 beamline personnel is gratefully appreciated. This research was supported by a scholarship from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Fonds Québécois de la Recherche sur la Nature et les Technologies (FQRNT) to LC and research grants from the Canadian Institutes of Health Research to JS and from NSERC and FQRNT to both NB and JS.

2.7 References

- Desveaux, D., Allard, J., Brisson, N. & Sygusch, J. (2002a). *Acta Cryst.* D58, 296-298.
- Desveaux, D., Allard, J., Brisson, N. & Sygusch, J. (2002b). *Nat Struct. Biol.* 9, 512-517.
- Desveaux, D., Després, C., Joyeux, A., Subramaniam, R. & Brisson, N. (2000). *Plant Cell*, 12, 1477-1489.
- Desveaux, D., Maréchal, A. & Brisson, N. (2005). *Trends Plant Sci.* 10, 95-102.
- Desveaux, D., Subramaniam, R., Després, C., Mess, J. N., Lévesque, C., Fobert, P. R., Dangel, J. L. & Brisson, N. (2004). *Dev. Cell*, 6, 229-240.
- Luft, J. R., Collins, R. J., Fehrman, N. A., Lauricella, A. M., Veatch, C. K. & DeTitta, G. T. (2003). *J. Struct. Biol.* 142, 170-179.
- Maréchal, A., Parent, J. S., Sabar, M., Veronneau-Lafortune, F., Abou-Rached, C. & Brisson, N. (2008). *BMC Plant Biol.* 8, 42.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* 276, 307-326.
- Vermel, M., Guermann, B., Delage, L., Grienemberger, J. M., Maréchal-Drouard, L. & Gualberto, J. M. (2002). *Proc. Natl. Acad. Sci. USA*, 99, 5866-5871.
- Yoo, H. H., Kwon, C., Lee, M. M. & Chung, I. K. (2007). *Plant J.* 49, 442-451.

Table 2-I. Diffraction data statistics

Dataset	St-WHY2 in free form	St-WHY2-ERE ₃₂
Beamline	X25, NSLS	X29, NSLS
Wavelength (Å)	1.08	1.08
Spacegroup	<i>F</i> 432	<i>F</i> 432
Unit-cell parameter (Å)	<i>a</i> = 164.58 Å	<i>a</i> = 167.16 Å
Resolution (Å)	50–2.20 (2.28–2.20)	50–2.20 (2.28–2.20)
No. of unique reflections	10208	10665
Redundancy	37.3 (19.5)	35.9 (9.9)
Completeness (%)	100.0 (99.9)	99.8 (98.7)
R_{sym}^{\dagger}	0.076 (0.92)	0.051 (0.56)
<i>I</i> /σ (<i>I</i>)	16.9 (3.8)	23.1 (3.8)

$\dagger R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of an individual measurement of the reflection with Miller indices hkl and $\langle I(hkl) \rangle$ is the mean intensity of that reflection.

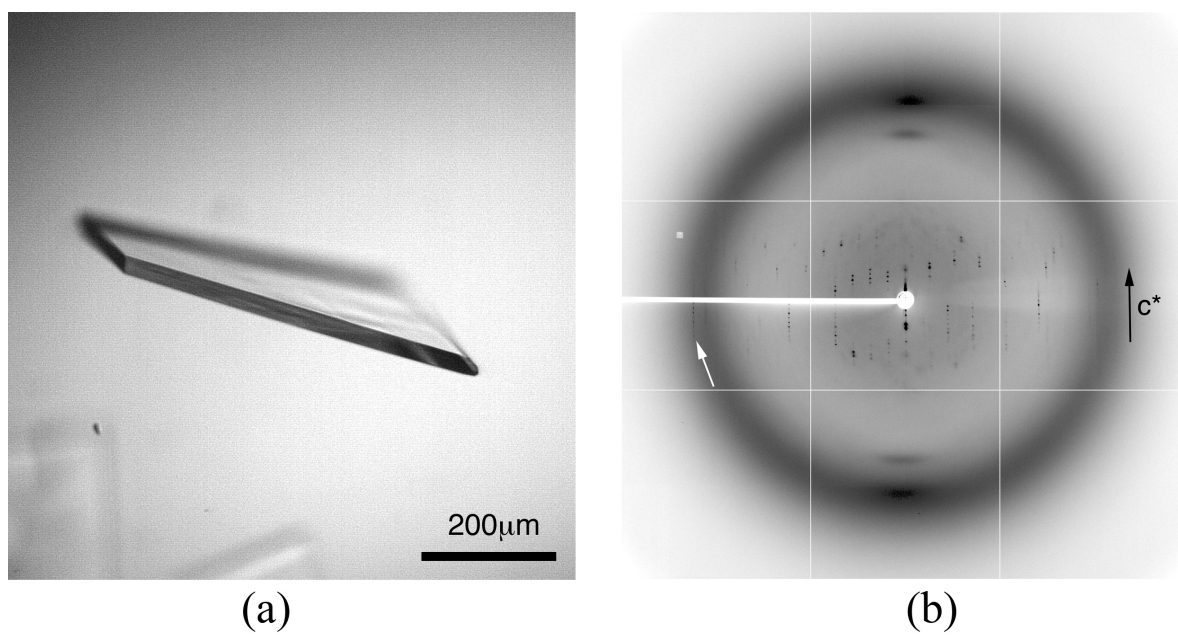


Figure 2-1. (a) Crystals of St-WHY1-ERE₃₂. (b) Diffraction pattern of St-WHY1-ERE₃₂ showing high anisotropy along the c* direction (black arrow). The white arrow highlights diffraction at 3.2 Å resolution.

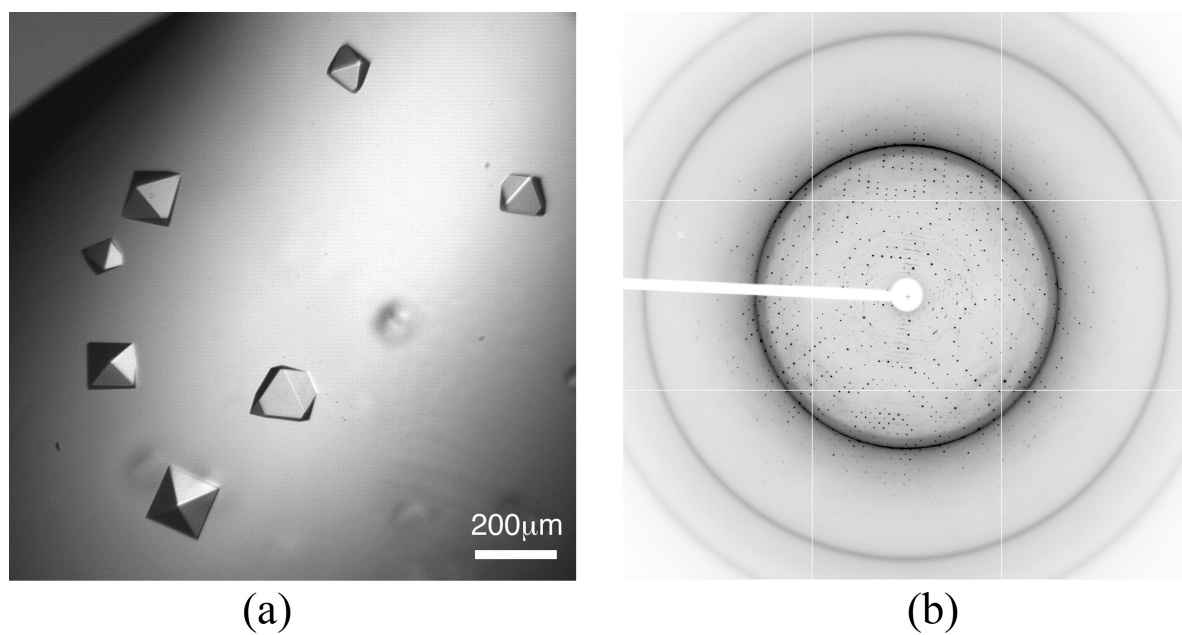


Figure 2-2. (a) Crystals of St-WHY2-ERE₃₂. (b) Diffraction pattern of St-WHY2-ERE₃₂ exhibiting low anisotropy.

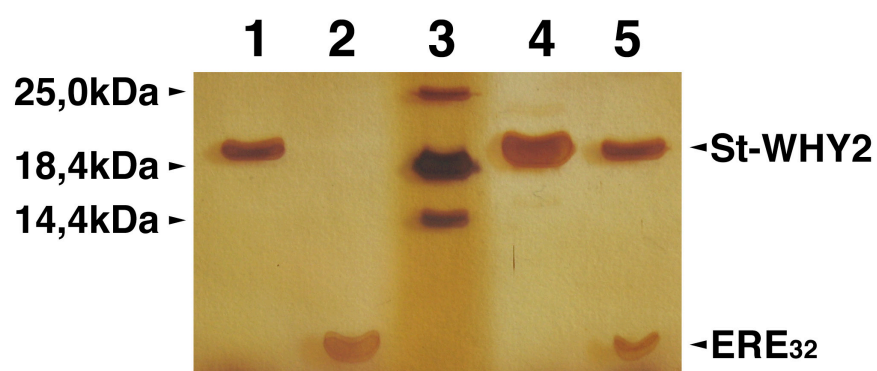


Figure 2-3. SDS-PAGE analysis of dissolved crystals. Lane 1, purified St-WHY2; lane 2, ERE₃₂; lane 3, molecular-weight markers (kDa); lane 4, washed and dissolved crystals of St-WHY2 in the free form; lane 5, washed and dissolved crystals of St-WHY2-ERE₃₂. Proteins and DNA were stained with silver nitrate.

CHAPITRE 3

Structures cristallines de complexes Whirly-ADN et leur rôle dans la réparation du génome des organites d'*Arabidopsis*

Article 2 : Crystal Structures of DNA-Whirly Complexes and Their Role in *Arabidopsis*
Organelle Genome Repair

Reprinted with permission from *the American Society of Plant Biologists*
(www.plantcell.org)

Crystal Structures of DNA-Whirly Complexes and Their Role in *Arabidopsis* Organelle Genome Repair

Laurent Cappadocia, Alexandre Maréchal, Jean-Sébastien Parent, Étienne Lepage, Jorgen Sygusch, and Normand Brisson

Department of Biochemistry
Université de Montréal
PO Box 6128, Station Centre-Ville
Montréal, Québec
H3C 3J7, Canada

The Plant Cell (2010) **22**: 1849-1867

3.1 Mise en contexte

Ce chapitre s'intéresse aux mécanismes de réparation de l'ADN dans les organites d'*Arabidopsis*, ainsi qu'au rôle des protéines Whirly dans ce processus. L'utilisation de la ciprofloxacine, un inhibiteur des ADN gyrases, a permis de produire des cassures bicaténares de l'ADN dans le génome des organites de plantes de type sauvage et dans des plantes n'exprimant plus les protéines Whirly. La réparation dépendante de courtes (de cinquante à mille paires de bases) séquences répétées d'ADN a été détectée par buvardage de type Southern. La réparation dépendante de très courtes (de cinq à cinquante paires de bases) séquences répétées d'ADN a quant à elle été détectée par PCR. Pour comprendre comment les protéines Whirly modulent la réparation des cassures bicaténares de l'ADN, nous avons utilisé une approche structurale en utilisant les données de diffraction des rayons X obtenues précédemment (cf. Chapitre 2). Nous avons ainsi montré que les protéines Whirly lient l'ADN avec peu de spécificité de séquence dans une conformation qui empêche la formation d'ADN bicaténaire.

Les résultats de ce chapitre ont été publiés dans la revue *The Plant Cell* en juin 2010. J'ai procédé à l'analyse des structures cristallines de WHY2 de pomme de terre en forme libre et en complexe avec de l'ADN. J'ai également effectué les expériences *in vitro* de caractérisation de la protéine WHY2. Alexandre Maréchal et Étienne Lepage ont analysé le phénotype visible de plantes d'*Arabidopsis* traitées avec les différents composés chimiques. Jean-Sébastien Parent a obtenu les plantes mutantes et a analysé le phénotype moléculaire de ces plantes par PCR. J'ai travaillé en collaboration avec Alexandre Maréchal pour cloner et séquencer les produits de PCR et j'ai effectué les expériences de buvardage de type Southern permettant la détection de la recombinaison homologue dans les mitochondries d'*Arabidopsis*. J'ai rédigé le présent manuscrit qui a par la suite été corrigé par tous les auteurs.

3.2 Abstract

DNA double-strand breaks are highly detrimental to all organisms and need to be quickly and accurately repaired. Although several proteins are known to maintain plastid and

mitochondrial genome stability in plants, little is known about the mechanisms of DNA repair in these organelles and the roles of specific proteins. Here, using ciprofloxacin as a DNA damaging agent specific to the organelles, we show that plastids and mitochondria can repair DNA double-strand breaks through an error-prone pathway similar to the microhomology-mediated break-induced replication observed in humans, yeast, and bacteria. This pathway is negatively regulated by the single-stranded DNA (ssDNA) binding proteins from the Whirly family, thus indicating that these proteins could contribute to the accurate repair of plant organelle genomes. To understand the role of Whirly proteins in this process, we solved the crystal structures of several Whirly-DNA complexes. These reveal a nonsequence-specific ssDNA binding mechanism in which DNA is stabilized between domains of adjacent subunits and rendered unavailable for duplex formation and/or protein interactions. Our results suggest a model in which the binding of Whirly proteins to ssDNA would favor accurate repair of DNA double-strand breaks over an error-prone microhomology-mediated break-induced replication repair pathway.

3.3 Introduction

Plant cells must ensure the accurate replication and faithful transmission of three different genomes located in the nucleus, mitochondria, and plastids. Although the maintenance of nuclear genome stability in eukaryotes is the subject of an intense research effort, the systems that prevent DNA lesions and rearrangements in organelle genomes remain poorly understood. This is unfortunate as the small genomes of plant organelles encode a variety of proteins with critical functions in oxidative phosphorylation or photosynthesis (Knoop, 2004; Raubeson and Jansen, 2005) and any substantial alteration in their DNA could have devastating consequences for plant growth and development (Newton et al., 2004).

Despite the fact that plant organelle genomes are constantly exposed to reactive by-products of the electron transport machineries, they are surprisingly more stable than the nuclear genome at the nucleotide level (Wolfe et al., 1987; Drouin et al., 2008). Contrasting with their slow rate of nucleotide substitution, plant organelle genome organization is complex and highly dynamic. Indeed, DNA rearrangements resulting from spurious

recombination between repeated sequences occur frequently in mitochondria and plastids (André et al., 1992; Maul et al., 2002). The prevention and repair of mutations as well as the regulation of recombination events require extensive genome maintenance systems. Plant organelle genomes encode very few, if any, of the proteins involved in these systems (Mackenzie, 2005; Day and Madesis, 2007). Their maintenance is thus mainly ensured by nuclear-encoded DNA replication, recombination, and repair proteins that are targeted to plastids and/or mitochondria. In recent years, some of these proteins have been identified, including the *Arabidopsis thaliana* and *Physcomitrella patens* RecA homologs (Shedge et al., 2007; Odahara et al., 2009), the *Arabidopsis* MutS-like MSH1 (Abdelnoor et al., 2003), the *Arabidopsis* organelle single-stranded DNA binding protein OSB1 (Zaegel et al., 2006), and the plastid-targeted Whirlies of *Arabidopsis* and maize (*Zea mays*; Maréchal et al., 2009).

Whirlies form a small family of proteins present mainly in the plant kingdom (Desveaux et al., 2005). Different members of this family can localize to the nucleus, plastids, or mitochondria (Desveaux et al., 2005; Krause et al., 2005; Grabowski et al., 2008; Maréchal et al., 2008, 2009; Prikryl et al., 2008). Whirly proteins preferentially bind single-stranded DNA (ssDNA; Desveaux et al., 2000, 2002, 2005) and perform numerous activities related to DNA metabolism, including the regulation of transcription (Desveaux et al., 2000, 2004; Xiong et al., 2009) and modulation of telomere length (Yoo et al., 2007). Interestingly, the maize Whirly protein WHY1 interacts with DNA from throughout the plastid genome but also binds to a subset of plastid RNAs and participates in chloroplast RNA metabolism (Prikryl et al., 2008). Resolution of the crystallographic structure of the *Solanum tuberosum* WHY1 protein revealed that Whirly proteins form tetramers resembling whirligigs. However, the precise mode of single-stranded nucleic acid binding, as well as the sequence specificity of Whirlies, has so far remained elusive (Desveaux et al., 2002).

The nuclear genome of *Arabidopsis* encodes three Whirly proteins: WHY1 and WHY3, which are targeted to plastids, and WHY2, which localizes to mitochondria (Krause et al., 2005; Maréchal et al., 2008, 2009). Recently, we demonstrated that in the absence of WHY1 and WHY3, the plastid genome of *Arabidopsis* becomes unstable and eventually accumulates rearranged DNA molecules causing chloroplast defects (Maréchal et al.,

2009). This phenomenon is not unique to *Arabidopsis*, as rearranged DNA molecules also accumulate in the plastids of maize lines with reduced levels of Zm-WHY1 (Maréchal et al., 2009). The rearranged molecules correspond to deletion, duplication, and/or circularization events that are mediated by DNA microhomologies (Maréchal et al., 2009). Interestingly, microhomology-mediated DNA rearrangements (MHMRs) are also present in the nuclear genomes of humans, yeast, and bacteria (Hastings et al., 2009a). In humans, these MHMRs are associated with numerous copy number variants, as well as with several diseases (Hastings et al., 2009a, 2009b; Lupski, 2009; Conrad et al., 2010), and appear following DNA-related stresses, notably in response to double-strand breaks (DSBs; Hastings et al., 2009a).

The similarity of the DNA rearrangements observed in plastids of Whirly mutants and in other species led us to suspect that the plastid genome rearrangements could be due to the error-prone repair of DNA following DNA damage and that the role of Whirly proteins in organelles could be to promote the accurate repair of DNA. We tested this hypothesis by treating wild-type and Whirly mutant plants with ciprofloxacin, a DSB-inducing genotoxic agent that targets the enzymes gyrases, which are localized in plant organelles (Wall et al., 2004). Our results indicate that DSBs lead to an accumulation of DNA rearrangements, including MHMRs, in both plastid and mitochondrial genomes. Importantly, plants lacking plastid or mitochondrial Whirlies accumulate higher levels of MHMRs than do the wild-type controls, suggesting that Whirlies might be components of the organelle repair machinery. To gain insights into the mechanisms by which Whirlies prevent DSB-induced rearrangements, we obtained several crystal structures of a mitochondrial Whirly protein in complex with ssDNA oligonucleotides. These structures reveal how Whirlies bind and protect the single-stranded form of DNA in a nonsequence-specific fashion. We propose a model in which Whirlies help maintain genome integrity under DNA stresses by binding to ssDNA regions of the genome and favoring conservative over error-prone repair pathways.

3.4 Results

3.4.1 DSB Induces MHMR Accumulation in the Plastids of *Arabidopsis*

To test the hypothesis that Whirlies are important for the accurate repair of DNA lesions, Columbia-0 (Col-0; wild type) and a *why1 why3* double knockout (formerly called KO1/3; Maréchal et al., 2009) line of *Arabidopsis* were treated with the quinolone ciprofloxacin. Quinolones are inhibitors of organelle-localized DNA gyrases (Wall et al., 2004) and induce DSBs in DNA (Gellert et al., 1977; Sugino et al., 1977; Snyder and Drlica, 1979; Chen et al., 1996; Rowan et al., 2010). The effect of ciprofloxacin on *Arabidopsis* wild-type plant survival was monitored and a LD₅₀ (dose required to kill half the members of the tested population) of 0.5 μ M ciprofloxacin was calculated (see Figure 3-S1). The error-prone repair of plastid DNA was assayed by monitoring the formation of DNA rearrangements using the PCR approach illustrated in Figure 3-1 and primer pairs scattered throughout the plastid genome (Maréchal et al., 2009).

As expected, treatment of *Arabidopsis* with ciprofloxacin induces leaf etiolation/variegation and inhibits seedling growth (Figure 3-2A; Wall et al., 2004). At 0.25 μ M ciprofloxacin, *Arabidopsis* wild-type plants displayed only slight growth delay and only 9% of the plants displayed etiolation/variegation (Figure 3-2B). At this same concentration, *why1 why3* plants displayed greater sensitivity to ciprofloxacin, with 88% of *why1 why3* plants displaying etiolation/variegation (Figure 3-2B). At 0.75 μ M ciprofloxacin, both genotypes were seriously affected and true leaves were unable to develop. We then monitored the DNA rearrangements in the plastid genomes of these plants using the PCR approach described in Figure 3-1. Figure 3-2C shows that whereas DNA rearrangements were almost undetectable in untreated wild-type plants, they were abundant and diverse in untreated *why1 why3* plants, in agreement with our previous findings (Maréchal et al., 2009). Treatment of plants with 0.25 μ M ciprofloxacin led to a stronger increase in the number of DNA rearrangements in *why1 why3* (13 additional PCR products) compared with wild-type plants (four additional PCR products; Figure 3-2D). At 0.75 μ M ciprofloxacin, both genotypes accumulated high levels of rearranged DNA molecules (Figure 3-2D).

To ensure that the PCR product accumulation was due to DSB induction rather than DNA gyrase inhibition, wild-type and *why1 why3* plants were treated with novobiocin, a DNA gyrase inhibitor that also targets organelle-localized DNA gyrases (Wall et al., 2004) but does not produce DSBs (Gellert et al., 1976, 1977; Sugino et al., 1977). Figures 3-2E and 3-2F show that novobiocin inhibited seedling growth to the same extent in both wild-type and *why1 why3* lines, with very little effect on leaf etiolation/variegation. However, no increase in the diversity or abundance of DNA rearranged molecules could be observed upon treatment with 100 μ M novobiocin, the highest concentration tested, in either wild-type or *why1 why3* plants (Figures 3-2G and 3-2H). This confirmed that DSB induction, rather than DNA gyrase inhibition, leads to DNA rearrangements in ciprofloxacin-treated plants.

3.4.2 DSBs Induce MHMR Accumulation in the Mitochondria of *Arabidopsis*

In contrast with *Arabidopsis* WHY1 and WHY3, no function in DNA metabolism has yet been ascribed to the mitochondria-localized protein WHY2. To verify if this protein has a role in maintaining mitochondrial genome stability under DNA stress conditions, we repeated the experiment with gyrase inhibitors on *why2* knockout plants. No phenotypic differences could be observed between *why2-1* (formerly called KO2; Maréchal et al., 2008) and wild-type plants in the presence or absence of ciprofloxacin (Figures 3-3A and 3-3B). We then monitored rearrangements in mitochondrial DNA following ciprofloxacin treatment using the PCR approach described above. Figures 3-3C and 3-3D show that rearranged DNA molecules are not abundant in wild-type or *why2-1* plants at 0 μ M ciprofloxacin concentration. However, treatment with 0.25 μ M ciprofloxacin led to a large increase in rearranged products for *why2-1* plants (18 additional PCR products) compared with wild-type plants (one additional PCR product; Figure 3-2D). At 0.75 μ M ciprofloxacin, rearranged DNA molecules accumulate to high levels in both wild-type and *why2-1* plants.

Figures 3-3E and 3-3F show that *why2-1* and wild-type plants were similarly affected by novobiocin. No rearranged DNA molecules could be detected in plants treated with either 0 or 100 μ M novobiocin (Figures 3-3G and 3-3H), thus confirming that mitochondrial DNA rearrangements in ciprofloxacin-treated plants depend on DSB induction. These results

indicate that, upon induction of DSBs, *why2-1* plants accumulate DNA rearrangements in the mitochondrial genome more readily than do wild-type plants.

3.4.3 DNA Rearrangements Are Mediated by Microhomologies

Cloning and sequencing of PCR products such as those shown in Figures 3-2C, 3-2G, 3-3C, and 3-3G led to the identification of 191 DNA rearrangement events (see Table 3-SI), some of which are presented in Table 3-I. The DNA rearrangements were classified in two groups: MHMR products that had at least 5 bp of microhomology at the junction and nonhomologous end joining (NHEJ)-like products that had <5 bp of microhomology at the junction (McVey and Lee, 2008). MHMR and NHEJ-like products constituted 83 and 17%, respectively, of all DNA rearrangements and were present in both deletion and duplication/circularization events (see Table 3-SI). Although MHMR junctions varied in sequence, they were frequently A/T rich and often contained homopolymeric tracts (Table 3-I). Some MHMR junctions appeared recurrently in different genotypes and/or in different plant treatments (see Table 3-SII) and were counted as independent events. This result suggests that DNA rearrangement hot spots exist. However, these may not all represent biologically relevant hot spots as DSBs could have a nonrandom distribution due to the binding of DNA gyrases at preferred locations (Fisher et al., 1981). The 49 mitochondrial MHMR junctions had a mean size of 12.6 ± 5.2 bp, whereas the 109 plastid MHMR junctions had a mean size of 14.5 ± 5.5 bp. Globally, the size of the MHMR junctions varied from 5 to 37 bp. No major difference was observed between the mean size of plastid junctions from ciprofloxacin-treated and untreated plants and that of mitochondrial junctions from ciprofloxacin-treated plants. This, plus the fact that MHMRs were more abundant in Whirly-deficient than in wild-type organelles (Table 3-SIII), indicates that a microhomology-dependent repair pathway is common to both organelles and is more active in Whirly-deficient plants. Importantly, rearrangements in the plastid genome of *why2-1* mutant plants and in the mitochondria of *why1 why3* mutant plants were at wild-type levels (see Figure 3-S2), thereby demonstrating the specificity of Whirly proteins for their respective organelle. Altogether, our data suggest that Whirlies could prevent break-induced DNA rearrangements of *Arabidopsis* organelle genomes.

3.4.4 DNA Rearrangements Mediated by Microhomologies Are Present as Substoichiometric Molecules in the Mitochondrial and Plastid Genomes

DNA gel blot analysis was performed on HindIII-digested DNA from wild-type and *why1 why3* plants using isolated PCR products as probes (see Figure 3-S3 and Table 3-SIV). In untreated plants, only the bands corresponding to the wild-type genome were visible in wild-type or *why1 why3* plants. Additional bands were detected in wild-type plants treated with 0.75 μ M ciprofloxacin and in *why1 why3* plants treated with 0.25 or 0.75 μ M ciprofloxacin. The fact that the additional bands always have a lower intensity than the band corresponding to the main genome suggests that they are present in substoichiometric amounts compared with the main genome. The DNA gel blot analysis of undigested DNA (see Figure 3-S3 and Table 3-SIV) revealed the existence of subgenomic-length molecules in wild-type and *why1 why3* plants treated with 0.75 μ M ciprofloxacin.

A similar analysis performed on *why2-1* and wild-type plants (see Figure 3-S4 and Table 3-SIV) indicated that the mitochondria of ciprofloxacin-treated plants also contain additional DNA molecules that are present at substoichiometric levels. Analysis of undigested DNA (see Figure 3-S3 and Table 3-SIV) further revealed the existence of subgenomic-length molecules in mitochondria. Together with our previous observations, these results suggest that substoichiometric DNA rearrangements accumulate in wild-type, *why2-1*, and *why1 why3* genotypes following ciprofloxacin treatment and that some of these rearrangements exist as subgenomic-length molecules.

3.4.5 Mitochondrial Recombination Mediated by Short Repeats (50 to 1,000 Nucleotides) Is Not Affected in Whirly Mutants

We next investigated the effect of gyrase inhibitors and Whirly mutants on recombination mediated by short repeats (50 to 1,000 nucleotides). This type of recombination is often detected in plant mitochondria (Mackenzie, 2005; Maréchal and Brisson, 2010; Woloszynska, 2010) and is modulated by a variety of proteins involved in maintaining mitochondrial genome stability, such as OSB1 (Zaegel et al., 2006), RecA3

(Shedge et al., 2007), RecA1 (Odahara et al., 2009), and Msh1 (Abdelnoor et al., 2003). DNA from wild-type, *why2-1*, and *why1 why3* plants were analyzed by DNA gel blot and probed with mitochondrial repeats as described by Arrieta-Montiel et al. (2009). DNA gel blot analyses revealed that both ciprofloxacin (see Figure 3-S5) and novobiocin (see Figure 3-S6) trigger mitochondrial recombination mediated by short repeated sequences. However, different recombination patterns were obtained for ciprofloxacin- and novobiocin-treated plants. This was expected since inhibition of DNA gyrase by these compounds leads to different outcomes. Indeed, ciprofloxacin triggers recombination-dependent repair of DSB (López and Blázquez, 2009), whereas novobiocin triggers recombination-dependent restart of stalled replication forks (Woelfle et al., 1993). Interestingly, the recombination patterns were similar for all plant genotypes. This suggests that Whirlies do not modulate the mitochondrial recombination mediated by short repeated sequences (50 to 1,000 nucleotides).

3.4.6 The DNA Rearrangements Are Induced Mainly by Stresses Targeting the Organelles

We next treated wild-type, *why2-1*, and *why1 why3* plants with bleomycin and hydroxyurea. In *Arabidopsis*, bleomycin triggers a DSB stress response (Menke et al., 2001; Tamura et al., 2002; Chen et al., 2008), whereas hydroxyurea induces a replication stress response (Ferreira et al., 1994; Roa et al., 2009). Treating wild-type and Whirly mutant plants with bleomycin (see Figure 3-S7) or hydroxyurea (see Figure 3-S8) did not result in an increase in plastid or mitochondrial MHMR. Furthermore, no increase in mitochondrial recombination mediated by short repeated sequences was observed upon treatment of *Arabidopsis* plants with these compounds (see Figure 3-S9). These results suggest that bleomycin and hydroxyurea have a weaker effect than do ciprofloxacin and novobiocin on the formation of DNA rearrangements in *Arabidopsis* organelle genomes. As shown in other species (Baugnet-Mahieu et al., 1971; Shen et al., 1995; Morel et al., 2008), this could be due to the fact that bleomycin and hydroxyurea target both the nucleus and the organelles, whereas compounds such as ciprofloxacin and novobiocin target mainly the organelles (Ye and Sayre, 1990). However, we cannot exclude the possibility that DNA damage induced by bleomycin and hydroxyurea in the organelles is repaired by alternative mechanisms.

3.4.7 The Crystal Structure of St-Why2 Reveals How ssDNA Is Bound

To understand how Whirly proteins prevent MHMR accumulation in the organelle genomes of *Arabidopsis*, we sought to determine how these proteins bind ssDNA. We thus solved the crystal structure of St-WHY2₄₈₋₂₁₆, a close homolog of At-WHY2 that also localizes to mitochondria (Vermel et al., 2002), in the free form and bound to ssDNA (Cappadocia et al., 2008). The structures were obtained at 2.2- to 2.7-Å resolution (Table 3-II). The St-WHY2₄₈₋₂₁₆ construct, hereafter called WHY2, encompasses the entire ssDNA binding domain but lacks the mitochondrial targeting sequence and the acidic/aromatic C-terminus (Figure 3-4A). These regions, which are not required for high affinity binding to the ssDNA (Cappadocia et al., 2008), are predicted to be disordered in solution. For the ssDNA sequence, we chose ERE₃₂, a 32-nucleotide ssDNA derived from the elicitor response element (ERE) because the interaction between this sequence and plant Whirlies is well documented (Desveaux et al., 2000, 2004). WHY2 and WHY2-ERE₃₂ crystallized in different chemical environments, yet they both gave rise to isomorphous crystals of the F432 space group. In the two crystals, WHY2 tetramers are disposed around fourfold crystallographic axes (Figures 3-4B and 3-4C). In the WHY2-ERE₃₂ structure, electron density corresponding to ssDNA was clearly visible (Figure 3-4C), enabling us to build a nine-nucleotide model. Although we could observe electron density for most nucleobases, we were unable to assign unambiguously the ERE₃₂ sequence into the electron density. We thus modeled the thymine-rich ERE₃₂ oligonucleotide as an all-thymine oligonucleotide.

Shown in Figure 3-4C is the electron density corresponding to ssDNA that was observed primarily on the edges and in between the β -sheets of adjacent protomers. Such a positioning of the DNA was unexpected. Indeed, while most single-stranded nucleic acid proteins use the core of their β -sheets as a primary binding platform (Horvath, 2008), WHY2-ssDNA interaction relies mainly on the binding of the DNA between properly positioned domains. As a consequence, binding to ssDNA exploits, but also depends on, the fourfold symmetry of the Whirly protein. The root mean square deviation (RMSD) of merely 0.5 Å for all C atoms between the free and the DNA-bound forms of WHY2 indicates that the protein does not undergo major conformational changes upon DNA binding. This is consistent with a need for properly positioned ssDNA binding residues.

Figure 3-S10 shows that whereas good electron density corresponding to ssDNA is observed near the edges of the β -sheets, this density fades out on the top of the β -sheets before reappearing near the other edge of the β -sheets. This suggests that the ssDNA is disordered and/or adopts multiple conformations on the top of the β -sheets. We propose that, in this region, long DNA molecules, such as the ERE₃₂, either thread from one ssDNA binding site to the other or enter/leave the complex.

3.4.8 WHY2 Binds ssDNA through a Conserved Mechanism

Figure 3-5A indicates that ssDNA is maintained in an extended conformation in each ssDNA binding site. The single-strand preference of WHY2 arises from steric impediment for the binding of a second strand at several positions on the binding surface. The presence of abrupt twists in the DNA backbone, such as between nucleotides 2 and 3, also argues against binding of a double-stranded DNA (dsDNA) helix. The mode of ssDNA binding is dominated by stacking and hydrophobic interactions between adjacent nucleobases and between nucleobases and aromatic/hydrophobic protein residues, consistent with the salt-resistant binding of St-WHY2 to ssDNA (Vermel et al., 2002). The polar contribution to DNA binding involves seven hydrogen bonds plus one water-mediated interaction (Figure 3-5A). Few sequence-specific interactions between St-WHY2 and ERE₃₂ were observed (see below). Indeed, most of the nucleobases have their edges, containing the sequence-specific binding moieties, exposed to the solvent, whereas the faces of the nucleobases make intimate contact with residues of the protein surface.

A structure-based sequence alignment reveals that most of the ssDNA binding residues are conserved among different plastid- and mitochondria-targeted Whirlies (Figure 3-5B). Structural alignment of St-WHY1 (PDB 1L3A) with St-WHY2 also indicates a strong structure conservation both at the monomeric level with an RMSD of 1.4 Å for 159 matched C atoms and at the tetrameric level with an RMSD of 1.9 Å for 642 matched C atoms. Furthermore, structural superposition of St-WHY1 with St-WHY2 reveals that the key ssDNA binding residues are properly positioned in St-WHY1 to contact ssDNA. This supports the hypothesis that plastid- and mitochondria-targeted Whirlies have closely related ssDNA binding interfaces and is consistent with a common role of Whirly proteins in both

organelles. Importantly, ssDNA binding residues are also conserved in *Arabidopsis* proteins, suggesting that the mechanism of ssDNA binding of St-WHY2 is similar to the ones of *Arabidopsis* WHY1, WHY2, and WHY3 (Figure 3-5B).

3.4.9 WHY2 Binds Nucleobases That Differ in Size and in Functional Groups

The local disorder of the electron density around the nucleobases at their junction with the β -sheets, as well as the scarcity of sequence-specific contacts between St-WHY2 and ERE₃₂, suggests that WHY2 could bind other ssDNA sequences with high affinity. We thus assayed the binding of St-WHY2 to different A/T-rich ssDNA, as plant organelle genomes are generally A/T rich (Fauron et al., 2004; Ravi et al., 2008). Quantitative electrophoretic mobility shift assays (EMSAs) revealed that, despite their different nucleotide sequences, the oligonucleotides ERE₃₂, cERE₃₂, rcERE₃₂, and dT₃₂ all interacted with WHY2 with similar nanomolar affinities (Figure 3-6A, Table 3-III). To gain structural insights on how WHY2 could accommodate these DNA sequences, we obtained the crystal structure of these oligonucleotides bound to St-WHY2 (Table 3-II) and compared them to that of WHY2-ERE₃₂. Comparison of the unbiased difference electron densities of the four oligonucleotides reveals that they adopt similar conformations in the ssDNA binding site (Figure 3-6B), suggesting that they interact with WHY2 through a common mechanism. The structures also reveal how WHY2 can accommodate bases that differ in size and in functional groups. Indeed, Ser-62 and Asp-142, which make sequence-specific interactions with T7 and T3 in the WHY2-dT₃₂ and WHY2-ERE₃₂ structures, respectively, make equivalent interactions with A7 and A3 in the structures of WHY2-cERE₃₂ and WHY2-rcERE₃₂ (Figures 3-6C and 3-6D). These compensating interactions demonstrate the capacity of WHY2 to bind different DNA sequences with high affinity.

3.4.10 WHY2 Binds to Single-Stranded Overhangs, Melts dsDNA, and Protects ssDNA

An important step in the repair of DSBs is the rapid resection of DNA by exonucleases to generate DNA with single-stranded overhangs (Persky and Lovett, 2008).

Because the choice of a repair pathway depends on the identity of the repair proteins that bind resected DNA ends, we verified that Whirly proteins could bind ssDNA overhangs. Figure 3-7A indicates that St-WHY2 can bind 16, but not 8, nucleotide-long 5' or 3' overhangs. Intriguingly, ssDNA molecules of the same size as these overhangs are not bound with high affinity by Whirlies (Yoo et al., 2007). This suggests either that the dsDNA moiety contributes to the high-affinity binding or that high-affinity binding is achieved through partially melting of the dsDNA moiety. As Whirlies were shown to have much less affinity for dsDNA than ssDNA (Desveaux et al., 2000; Prikryl et al., 2008), and since our structural data suggest that dsDNA is a poor substrate for Whirlies, we tested whether Whirlies could melt dsDNA to bind additional nucleotides. We incubated a radiolabeled DNA duplex with WHY2 and monitored the WHY2-induced denaturation of dsDNA by electrophoresis. Figure 3-7B shows that more ssDNA is produced as the concentration of WHY2 is increased. This indicates that WHY2 can destabilize DNA duplexes, probably by binding with a higher affinity to the single-stranded form of DNA and driving the equilibrium toward the formation of ssDNA. This could explain the capacity of WHY2 to bind short ssDNA overhangs with high affinity.

In addition to its ssDNA binding capacity, WHY2 effectively protects DNA against nuclease degradation. As shown in Figure 3-7C, phage M13mp18 ssDNA preincubated with a saturating amount of WHY2 is protected against digestion by mung bean (*Vigna radiata*) nuclease, whereas free M13mp18 ssDNA is readily digested. Also, Figure 3-7D shows that WHY2 can protect the single-stranded overhang of a duplex against degradation by the exonuclease activity of T4 DNA polymerase. The fact that WHY2 protects ssDNA against both mung bean nuclease and T4 DNA polymerase likely suggests that WHY2 can prevent these nucleases from accessing the ssDNA.

3.5 Discussion

3.5.1 Binding of Whirly Proteins to Nucleic Acids

There have been conflicting reports concerning the sequence specificity of Whirly proteins. Indeed, Whirlies were reported to bind three nuclear sequences that share little

sequence similarity with each other: an ERE element (Desveaux et al., 2000, 2004), an *Arabidopsis* telomeric repeat (Yoo et al., 2007), and a distal element upstream of the *Arabidopsis* kinesin gene KP1 (Xiong et al., 2009). In plant organelles, Whirly proteins were shown to bind many regions of the plastid or the mitochondrial genome without obvious sequence consensus in vivo (Maréchal et al., 2008, 2009; Prikryl et al., 2008). This work demonstrates that Whirlies can bind ssDNA molecules that differ in nucleotide sequence using a binding mechanism involving mainly contacts with nonsequence-specific moieties of the ssDNA and allowing Whirlies to establish equivalent interactions with different nucleobases. Because Whirlies fused to green fluorescent protein accumulate abundantly in the organelles, whereas they are difficult to detect in the nucleus (Krause et al., 2005), we propose that, in the nucleus, the low concentration of Whirly proteins would enable them to bind only a limited number of sites that have a pronounced single-stranded character (e.g., A/T-rich sequences of transcribed regions or telomeric repeats). In the organelles, the abundance of Whirly proteins and their limited sequence specificity would enable them to bind all available ssDNA.

In addition to ssDNA, Whirlies were shown to bind RNA in maize (Prikryl et al., 2008). Furthermore, a close structural neighbor of plant Whirlies also binds RNA (Schumacher et al., 2006). Our crystallographic models are consistent with a role of Whirly proteins in RNA metabolism. Indeed, modeling of ssRNA in place of ssDNA leads only to small clashes, thus suggesting that RNA could be accommodated in an ssDNA-like conformation (see Figure 3-S11). However, RNA appears to be bound less strongly than DNA because Whirlies were found to be associated with DNA from throughout the genome but with only a subset of RNA molecules in vivo (Prikryl et al., 2008).

3.5.2 DSBs Can Be Repaired through a Microhomology-Mediated Break-Induced Replication Mechanism in Plant Organelles

Plants treated with ciprofloxacin accumulate rearranged DNA molecules in both plastids and mitochondria. Most of the DNA rearrangements contain microhomologies at their junction, suggesting that at least part of the ciprofloxacin-induced DSBs are repaired through an error-prone microhomology-dependent pathway. We previously proposed that a

microhomology-mediated break-induced replication (MMBIR) pathway could explain the accumulation of MHMR in both untreated *why1 why3 Arabidopsis* and *Why1* knockdown maize plants (Maréchal et al., 2009). The MMBIR pathway relies on the microhomology-dependent restart of DNA synthesis on a different template following collapse of the replication fork (Hastings et al., 2009a). Microhomologies are sufficient to initiate DNA synthesis as certain DNA polymerases can efficiently use mismatched primers or primers as short as two to three base pairs as substrates (Cannistraro and Taylor, 2007; Hastings et al., 2009a). Our present finding that MHMR events that appear in untreated plants are increased upon ciprofloxacin treatment suggests that a MMBIR pathway is involved in the error-prone repair of DSBs in organelles.

The use of a MMBIR pathway is consistent with our observation of both duplications/circularization and deletion events, whereas pathways such as microhomology-mediated end joining or single-strand annealing can account only for deletion events (McVey and Lee, 2008). A NHEJ repair pathway could be involved in the formation of NHEJ-like events. However, there is little support for the existence of such a repair pathway in plant organelles, particularly in plastids (Odom et al., 2008; Kohl and Bock, 2009). Furthermore, whereas MHMRs are detected in the absence or presence of ciprofloxacin, all but one NHEJ-like event are detected in ciprofloxacin-treated plants, thus raising the possibility that these events could be by-products of the ciprofloxacin-mediated DNA gyrase inhibition (Marvo et al., 1983). The fact that mismatches are detected in 91 out of 158 MHMR junctions relative to one or both parental strands (see Table 3-SI) is consistent with the low-specificity requirements of the MMBIR pairing process. Finally, MMBIR can account for complex cases of DNA rearrangements (Hastings et al., 2009a), including the formation of molecules containing multiple MHMRs, which have also been observed in our study (Table 3-I). Our results also suggest that MHMR products exist at substoichiometric levels. These DNA variants could accumulate to higher levels through a process known as substoichiometric shifting, which is well documented for mitochondrial DNA rearrangements (Small et al., 1987; Janska et al., 1998; Maréchal and Brisson, 2010; Woloszynska, 2010). The fact that, in *why1 why3* plants, MHMR products present at low abundance in the chloroplasts of green sectors accumulate to high levels in the variegated sectors (Maréchal et al., 2009) suggests that this substoichiometric shifting could also occur in chloroplasts. A substoichiometric shift may

also facilitate the transmission of rearranged molecules. Indeed, in the *why1 why3* plants, variegation could readily be inherited provided that the female parent displayed a strong variegation phenotype (Maréchal et al., 2009). In this study, however, no increase in DNA rearrangements was observed in the progeny of plants treated with 0.25 μ M ciprofloxacin. It is thus possible that DNA rearrangements, when present at substoichiometric levels, have poor transmission efficiency. Alternatively, germ cells may be less sensitive to ciprofloxacin treatment than are somatic cells. Globally, the characteristics of the MHMR events obtained in this study support the idea that MMBIR acts as a DNA repair pathway in plant organelles (Figure 3-8).

3.5.3 Effects of DSBs and Whirly Mutants on Mitochondrial Recombination Mediated by Short Repeated Sequences (50 to 1,000 Nucleotides)

Plants treated with ciprofloxacin or novobiocin displayed an increase in mitochondrial recombination mediated by short repeated sequences. As the plastid genome of *Arabidopsis* does not contain short repeated sequences capable of supporting aberrant homologous recombination (Maréchal and Brisson, 2010), we were unable to determine if ciprofloxacin or novobiocin affected homologous recombination in plastids. Whirlies do not appear to modulate mitochondrial DNA recombination mediated by short repeated sequences. This was expected for the plastid-localized proteins WHY1 and WHY3 and is consistent with our previous observations that DNA recombination products mediated by short repeats do not accumulate in untreated *Arabidopsis* plants lacking the mitochondria-localized WHY2 (Maréchal et al., 2008). As At-WHY2 appears to modulate only the DSB-induced mitochondrial MMBIR pathway, it is possible that these repair responses are mechanistically distinct.

However, we cannot rule out a role for WHY2 in homologous recombination because other proteins may function redundantly with the Whirlies in this process. Indeed, many different ssDNA binding proteins are located in mitochondria, including the OSB proteins (Zaegel et al., 2006) and the mitochondrial ssDNA binding proteins (Edmondson et al., 2005).

3.5.4 Whirly Proteins Share Common Characteristics with Eubacterial ssDNA Binding Proteins

The eubacterial ssDNA binding proteins SSBs are critical players during DNA replication, recombination, and repair processes. Often envisioned as simple ssDNA-coating proteins, they actually work as organizational scaffolds that recruit genome maintenance complexes when and where they are most required (Cox, 2007; Shereda et al., 2008). These proteins share a number of characteristics with the Whirly proteins even though they are not structurally related: (1) they are both tetrameric proteins that bind ssDNA with high affinity and modest sequence preference, (2) they use mainly hydrophobic/aromatic residues to contact ssDNA and, as a result, their binding to ssDNA is salt resistant, (3) they can destabilize dsDNA, (4) they can prevent degradation of ssDNA by nucleases, and (5) they both possess a conserved C-terminal acidic/aromatic tail. Usually tethered to the protein surface, the acidic/aromatic C terminus of the eubacterial ssDNA binding proteins SSBs is believed to be freed and rendered available for interacting with target proteins upon ssDNA binding (Lu and Keck, 2008; Shereda et al., 2008, 2009). Interestingly, structural comparison between St-WHY1 and St-WHY2 revealed that the C terminus in WHY1 and the ssDNA in WHY2 both interact with the same residues in the β -sheets. This demonstrates that the C terminus and the ssDNA can compete for the same binding surface on the Whirly proteins and that their interaction with the Whirly domain is mutually exclusive. The competition between ssDNA and an acidic/aromatic C-terminal tail is consistent with our previous results showing that a St-WHY1 protein mutated at its C terminus has an increased affinity for ssDNA (Desveaux et al., 2005). Our results are thus consistent with a model in which, upon binding of Whirly proteins to ssDNA, the acidic-aromatic C-terminal tail would become available for interacting with target proteins in an SSB-like mode.

3.5.5 Possible Roles for Whirly Proteins in the Repair of DSBs

Whirlies could maintain genome stability by favoring the repair of DSBs by error-free homologous recombination and/or by restricting the DSB repair by error-prone MMBIR. Based on our results, we propose that Whirlies could bind and protect resected DNA ends in a sequence-independent manner at break sites (Figure 3-8). Upon binding to resected DNA

ends, Whirlies would prevent annealing of ssDNA overhangs to microhomologous sequences present in stretches of ssDNA elsewhere in the genome. Alternatively, or concomitantly, Whirlies could prevent MMBIR by binding to these ssDNA stretches, which may occur in many places in the genome, including in DNA secondary structures, in replication forks, in stalled transcription complexes, and in promoter regions (Hastings et al., 2009a). In addition, Whirlies might regulate the access of repair proteins to ssDNA by competing with them for binding to ssDNA, thereby influencing the choice of a repair pathway. Whirlies could even promote the accurate repair of DSBs by recruiting proteins involved in DNA repair through their acidic/aromatic C terminus in an SSB-like fashion.

No accumulation of rearranged DNA molecules could be detected in the mitochondria of *why2-1* plants in absence of ciprofloxacin treatment. This could suggest that mitochondrial DNA is less subject to genotoxic stresses under normal growth conditions or that WHY2 fulfils a different function than do WHY1 and WHY3 in *Arabidopsis*. A partial functional redundancy by the OSBs and mitochondrial SSBs could also explain why, in the absence of DNA stresses, *why2-1* plants do not show an increase in genome instability, contrary to what is observed in plastids of *Arabidopsis* and maize plants lacking plastid-localized Whirlies (Maréchal et al., 2009). This idea is consistent with the apparent lack of SSB homologs targeted to plastids. A functional redundancy could also explain our observation that, upon DSB induction, DNA rearrangements accumulate more readily in plastids of *why1 why3* plants than in mitochondria of *why2-1* plants. Further work is needed to determine if WHY2 works redundantly with other mitochondria-localized ssDNA binding proteins or if the protection against MHMR observed in *why2-1* mutants could be due to a more general role of WHY2 in DNA metabolism.

Interestingly, TIF1, a distant homolog of the plant Whirlies (Saha et al., 2001) safeguards chromosomes from DNA damage in both nuclei of the ciliate *Tetrahymena thermophyla*. TIF1-deficient cells are hypersensitive to genotoxic stress and are defective in the activation of the intra-S phase checkpoint mediated by the sensor/transducer kinase ATR (Yakisich et al., 2006). These studies suggest that a role of Whirlies in the maintenance of genome stability may be conserved between plants and ciliates.

Our finding that MHMR products accumulate at high concentrations of ciprofloxacin even in the presence of Whirly proteins suggests that accurate repair of DNA may be bypassed if severe DNA damage occurs. Interestingly, in other organisms, MHMR was detected mainly when canonical repair pathways such as NHEJ or homologous recombination were overloaded or challenged, thus leading to the idea that microhomology-mediated DNA repair acts as a backup DNA repair pathway (Wang et al., 2003). In *Arabidopsis* organelles, multiple DSBs may cause depletion of repair proteins due to their recruitment to multiple damage sites, allowing the accumulation of MHMRs. Our observation that MHMR products accumulate to high levels at 0.75 μ M ciprofloxacin both in the absence and the presence of Whirly proteins is consistent with this idea. However, the use of an error-prone repair pathway under DNA stress may be more than a backup pathway. Indeed, as MHMR enables DNA rearrangements including copy number variation (Hastings et al., 2009a), it promotes rapid changes in the genome. Some of these changes may confer selective advantages to plants under stress. A similar mechanism, termed stress-induced mutation, exists in bacteria and is also dependent on error-prone repair of DSBs (Ponder et al., 2005). The fact that certain plants maintain DNA rearrangements mediated by microhomologies in their organellar genome provides indirect support for their selective advantage (Ogihara et al., 1988; Kanno et al., 1993; Moeykens et al., 1995).

3.6 Methods

3.6.1 Plant Material and Growth Conditions

Sterilized seeds of *Arabidopsis thaliana* ecotype Col (wild type), the *why1 why3* double mutant, and the *why2* single mutant were germinated on Murashige and Skoog basal media (Sigma-Aldrich) containing ciprofloxacin or novobiocin at the indicated concentrations and collected after 3 weeks of growth under long-day conditions (16 h light/8 h dark). Bleomycin and hydroxyurea treatments were performed as described (Chen et al., 2008; Roa et al., 2009). The *why2-1* and *why1 why3* mutant lines (formerly called KO2 and KO1/3, respectively) were described previously (Maréchal et al., 2008, 2009). An additional *Arabidopsis* line lacking WHY2 (*why2-2*; SALK_016156) was tested for its sensitivity to the DNA damaging agents and the generation of MHMR products during ciprofloxacin

treatment (see Figure 3-S12C) and similar results were obtained. A protein gel blot was performed using α -WHY2 polyclonal antibodies (Maréchal et al., 2008) to verify that both *why2* lines are completely devoid of WHY2 protein (see Figure3-S12B).

3.6.2 Detection of DNA Rearrangements

DNA was isolated from plants using a cetyl trimethylammonium bromide DNA extraction protocol (Weigel and Glazebrook, 2002). PCR was conducted using the Taq polymerase (Genscript) according to the manufacturer's instructions. DNA rearrangement events were detected using both outward- and inward-facing PCR primers spaced by 5 to 30 kb. Thirteen and 10 PCR primer pairs scattered in the plastid genome were used for ciprofloxacin- and novobiocin-treated plants, respectively. Twenty and 10 PCR primer pairs scattered in the mitochondrial genome were used for ciprofloxacin- and novobiocin-treated plants, respectively. The sequences of the PCR primers are listed in Table 3-SV. PCR was performed on DNA samples from wild-type, *why2*, and *why1 why3* plants and analyzed by gel electrophoresis. All visible DNA bands were isolated and cloned into the pDrive vector (Qiagen). The DH5 strain was used for transformation, and the plasmids from two randomly picked colonies were sequenced. When both sequencing products were identical, only one product was considered, whereas when sequencing products differed, both products were considered. Sequencing of visible bands confirmed that >96% of all cloned PCR products were due to DNA rearrangements, with the remaining events being caused by nonspecific annealing of the oligonucleotides at undesired locations. DNA rearrangements are listed in Table 3-SI.

3.6.3 Analysis of DNA Rearrangements by DNA Gel Blots

To evaluate the relative abundance of MHMRs compared with normal organelle genomes, rearranged DNA molecules were amplified by PCR using the oligonucleotides listed in Table 3-SV. Cloning and sequencing of PCR products confirmed the occurrence of DNA rearrangements (see Table 3-SIV). PCR products were internally radiolabeled using [α - 32 P]dATP (6,000 Ci/mmol; Perkin-Elmer Life Science) and Klenow polymerase (Fermentas) and were used as probes. DNA gel blots were performed as described previously

(Maréchal et al., 2009). Detection of recombination mediated by short repeats (50 to 1,000 nucleotides) was performed as described (Arrieta-Montiel et al., 2009). Briefly, PCR primers listed in Table 3-SV were used to amplify three mitochondrial repeats. The repeats were internally radiolabeled using [α - 32 P]dCTP (3,000 Ci/mmol; Perkin-Elmer Life Science) and Klenow polymerase (Fermentas) and were used as probes.

3.6.4 St-WHY2 Cloning, Expression, and Purification

Details concerning cloning, expression, and purification of St-WHY2 are reported elsewhere (Cappadocia et al., 2008). Briefly, WHY248-216 was amplified from *Solanum tuberosum* total RNA and cloned into the plasmid vector pET-21a (Novagen), which encodes a C-terminal hexahistidine tag. St-WHY2 was expressed in *Escherichia coli* BL21 (DE3) strain with 1 mM isopropyl β -D-thiogalactopyranoside for 3 h at 37°C. The cells were lysed by alumina grinding and the lysate resuspended in 20 mM sodium phosphate buffer, pH 7.5, containing 500 mM NaCl and 25 mM imidazole. The recombinant protein was purified by applying the supernatant to a HiTrap Chelating column (GE Healthcare) followed by size exclusion chromatography on a Superdex 200 16/60 (GE Healthcare) in 20 mM Tris-HCl, pH 8.0, 100 mM NaCl.

3.6.5 Data Collection and Structure Determination

St-WHY2 in the free form and St-WHY2-ssDNA complexes were crystallized as previously described (Cappadocia et al., 2008). Briefly, protein-DNA complexes were crystallized by the hanging drop vapor diffusion method using a precipitant solution containing 100 mM Tris/HCl, pH 8.0, 13.5 to 25% (w/v) PEG6000, and 1 to 2 M LiCl. The free protein was crystallized by the hanging drop vapor diffusion method using a precipitant solution containing 100 mM MOPS, pH 7.0, 21 to 27% (w/v) PEG1000, and 100 to 400 mM $\text{NH}_4\text{H}_2\text{PO}_4$. Diffraction data were collected using ADSC Quantum 315 CCD detectors at beamlines X25 and X29 at the National Synchrotron Light Source (NSLS) at the Brookhaven National Laboratory or using a MAR-225 CCD detector at beamline 22-ID at the Advanced Photon Source at the Argonne National Laboratory. Initial phases for the St-WHY2-ERE₃₂ complex were obtained by molecular replacement using St-WHY1 (PDB 1L3A) as a search

template. Phases were improved by iterative cycles of model building with Coot (Emsley and Cowtan, 2004) and refinement with CNS (Brünger et al., 1998) and Phenix (Adams et al., 2002). DNA nucleotides were modeled one nucleobase at a time from difference electron density maps. The refined structure of St-WHY2-ERE₃₂ was used as a molecular replacement search template for structural determination of data sets collected from crystal of other St-WHY2-ssDNA complexes. Clear density was visible for protein residues 55 to 215 in St-WHY2. Since we could not attribute unambiguously the specific sequences of the ERE₃₂, rcERE₃₂, and cERE₃₂ oligonucleotides to the electron density, we assigned these oligonucleotides as all-thymine, all-adenine, and all-adenine oligonucleotides, respectively. The DNA occupancy in each structure was calculated from the temperature factors and it ranges from 0.7 to 0.8. Test data sets were randomly selected from the observed reflections prior to refinement. The figures were prepared using PyMOL (DeLano, 2002).

3.6.6 EMSAs

The apparent dissociation constants for St-WHY2 bound to DNA were determined by EMSA. The ssDNA oligonucleotides were obtained from Integrated DNA Technologies and radiolabeled at the 5' end using [γ -³²P]ATP (6,000 Ci/mmol; Perkin-Elmer Life Science) and T4 polynucleotide kinase (Fermentas) following the manufacturer's instructions. Binding reactions were done for 30 min at 20°C by incubating 15,000 cpm radiolabeled ssDNA (<100 pM) with the serially diluted St-WHY2 protein in a buffer containing 10 mM Tris-HCl, pH 8.0, 150 mM KCl, 0.5 mM EDTA, pH 8.0, and 15% (v/v) glycerol. Reactions were loaded on 10% (w/v) 29:1 acrylamide:bisacrylamide gels with 4.5 mM Tris, 4.5 mM boric acid, and 1 mM EDTA, pH 8.0, at 20°C under 150 V constant voltage for 20 min and then subjected to autoradiography. The autoradiograms were scanned and the intensity of the bands corresponding to bound and free radiolabeled ssDNA were quantified using ImageJ (NIH; <http://rsb.info.nih.gov/ij/>). The data was plotted and analyzed using Prism 5 Demo (GraphPad Software). The apparent dissociation constants were calculated by fitting the plot of the fraction of ssDNA bound versus protein concentration to the Hill equation.

The binding of St-WHY2 to ssDNA overhangs was also assessed by EMSA using DNA duplex formed by annealing two of the following oligonucleotides: YG, YC, YG8_3,

YC8_5, YG16_3, and YC16_5. The sequences of these oligonucleotides are shown in Table 3-SV.

3.6.7 DNA Destabilization Assay

Oligonucleotide R2 was radiolabeled at the 5' end using [γ - 32 P]ATP (6,000 Ci/mmol; Perkin-Elmer Life Science) and T4 polynucleotide kinase (Fermentas) following the manufacturer's instructions and annealed with oligonucleotide R1. The annealed duplex was incubated with St-WHY2 or BSA for 30 min at 20°C. The reactions were stopped by the addition of a 40-fold excess of unlabeled R2 and SDS to a final concentration of 0.2%. Reactions were separated on 7.5% (w/v) 29:1 acrylamide:bisacrylamide gels with 4.5 mM Tris, 4.5 mM boric acid, and 1 mM EDTA, pH 8.0, at 20°C under 130 V constant voltage for 30 min. The sequences of oligonucleotides R1 and R2 are shown in Table 3-SV.

3.6.8 Mung Bean Nuclease Protection Assay

Two hundred nanograms of M13mp18 ssDNA (USB) either alone or prebound for 30 min at 20°C with St-WHY2 at a 1:10 protein/nucleotide ratio were incubated with 1U of mung bean (*Vigna radiata*) nuclease (New England Biolabs) for 0, 15, 30, or 60 min at 30°C. Reactions were stopped by incubation on ice and addition of SDS and EDTA to final concentrations of 0.1% (w/v) and 1.5 mM, respectively. Reactions were loaded on 0.7% (w/v) agarose gels and migrated under 100 V of constant voltage.

3.6.9 T4 DNA Polymerase Protection Assay

ssDNA or a dsDNA with a 3' overhang of 16 nucleotides (3'-dsDNA) either alone or prebound for 30 min at 20°C with 1 μ M St-WHY2 was incubated with 0.02 units of T4 DNA Polymerase (T4; Fermentas) for 0, 15, 30, or 60 min at 20°C. Reactions were stopped by incubation on ice and addition of SDS to a final concentration of 0.1% (w/v). Reactions were separated on 10% (w/v) 29:1 acrylamide:bisacrylamide gels with 4.5 mM Tris, 4.5 mM boric acid, and 1 mM EDTA, pH 8.0, at 20°C under 150 V constant voltage for 20 min and then subjected to autoradiography.

3.6.10 Accession Numbers

Sequence data from this article can be found in the *Arabidopsis* Genome Initiative, GenBank/EMBL, or The Institute for Genomic Research database under the following accession numbers: St-WHY2 (HM234504), At-WHY1 (AT1G14410), At-WHY2 (AT1G71260), and At-WHY3 (AT2G02740). The accession numbers for the mitochondrial and plastid genomes are NC_001284 (Unsold et al., 1997) and NC_000932 (Sato et al., 1999), respectively. The atomic coordinates of St-WHY2 in the free form, St-WHY2-ERE₃₂, St-WHY2-dT₃₂, St-WHY2-cERE₃₂, and St-WHY2-rcERE₃₂, have been deposited in the Protein Data Bank (www.rcsb.org) under ID codes 3N1H, 3N1I, 3N1J, 3N1K, and 3N1L, respectively. The ID code for the crystal structure of St-WHY1 in the free form is 1L3A.

3.7 Acknowledgments

The assistance of Mathieu Coinçon, Sandra Grondin, and Bruno Piché is gratefully acknowledged. We thank James Omichinski for valuable discussions. We also thank the High Throughput Crystallization Service at the Hauptman-Woodward Institute, Buffalo, NY, for determining the initial crystallization conditions. Research carried out at the NSLS and at the APS was supported by the U.S. Department of Energy, Division of Materials Sciences and Division of Chemical Sciences, and by the U.S. Department of Energy, Office of Science, respectively. Assistance by X25, X29, and 22ID beamline personnel is gratefully appreciated. L.C. and A.M. were supported by fellowships from the Natural Sciences and Engineering Research Council of Canada (NSERC) and from the Fonds québécois de la recherche sur la nature et les technologies (FQRNT). This research was supported by grants from NSERC and FQRNT to both N.B. and J.S. and from the Canadian Institutes of Health Research to J.S.

3.8 References

- Abdelnoor, R. V., Yule, R., Elo, A., Christensen, A. C., Meyer-Gauen, G., and Mackenzie, S. A. (2003). Substoichiometric shifting in the plant mitochondrial genome is influenced by a gene homologous to MutS. *Proc. Natl. Acad. Sci. USA* 100: 5968–5973.
- Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W.,

- Read, R.J., Sacchettini, J.C., Sauter, N.K., and Terwilliger, T.C. (2002). PHENIX: Building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.* 58: 1948–1954.
- André, C., Levy, A.A., and Walbot, V. (1992). Small repeated sequences and the structure of plant mitochondrial genomes. *Trends Genet.* 8: 128–132.
- Arrieta-Montiel, M.P., Shedge, V., Davila, J., Christensen, A.C., and Mackenzie, S.A. (2009). Diversity of the *Arabidopsis* mitochondrial genome occurs via nuclear-controlled recombination activity. *Genetics* 183: 1261–1268.
- Baugnet-Mahieu, L., Goutier, R., and Baes, C. (1971). Differential response of mitochondrial and nuclear DNA syntheses to hydroxyurea in normal and regenerating rat liver. *Biochem. Pharmacol.* 20: 141–149.
- Brünger, A.T., et al. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* 54: 905–921.
- Cannistraro, V.J. and Taylor, J.S. (2007). Ability of polymerase eta and T7 DNA polymerase to bypass bulge structures. *J. Biol. Chem.* 282: 11188–11196.
- Cappadocia, L., Sygusch, J., and Brisson, N. (2008). Purification, crystallization and preliminary X-ray diffraction analysis of the Whirly domain of StWhy2 in complex with single-stranded DNA. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 64: 1056–1059.
- Chen, C.R., Malik, M., Snyder, M., and Drlica, K. (1996). DNA gyrase and topoisomerase IV on the bacterial chromosome: Quinolone-induced DNA cleavage. *J. Mol. Biol.* 258: 627–637.
- Chen, I.P., Mannuss, A., Orel, N., Heitzeberg, F., and Puchta, H. (2008). A homolog of ScRAD5 is involved in DNA repair and homologous recombination in *Arabidopsis*. *Plant Physiol.* 146: 1786–1796.
- Conrad, D.F., Bird, C., Blackburne, B., Lindsay, S., Mamanova, L., Lee, C., Turner, D.J., and Hurles, M.E. (2010). Mutation spectrum revealed by breakpoint sequencing of human germline CNVs. *Nat. Genet.* 42: 385–391.
- Cox, M.M. (2007). Regulation of bacterial RecA protein function. *Crit. Rev. Biochem. Mol. Biol.* 42: 41–63.

- Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., III, Snoeyink, J., Richardson, J.S., and Richardson, D.C. (2007). MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 35: W375–383.
- Day, A. and Madesis, P. (2007). DNA replication, recombination and repair in plastids. In *Cell and Molecular Biology of Plastids. Topics in Current Genetics*, Vol. 19, R. Bock, ed (Berlin, Heidelberg, Germany: Springer), pp. 65–119.
- DeLano, W.L. (2002). The Pymol Molecular Graphics System. (San Carlos, CA: DeLano Scientific).
- Desveaux, D., Allard, J., Brisson, N., and Sygusch, J. (2002). A new family of plant transcription factors displays a novel ssDNA-binding surface. *Nat. Struct. Biol.* 9: 512–517.
- Desveaux, D., Després, C., Joyeux, A., Subramaniam, R., and Brisson, N. (2000). PBF-2 is a novel single-stranded DNA binding factor implicated in *PR-10a* gene activation in potato. *Plant Cell* 12: 1477–1489.
- Desveaux, D., Maréchal, A., and Brisson, N. (2005). Whirly transcription factors: Defense gene regulation and beyond. *Trends Plant Sci.* 10: 95–102.
- Desveaux, D., Subramaniam, R., Després, C., Mess, J.N., Lévesque, C., Fobert, P.R., Dangl, J.L., and Brisson, N. (2004). A «Whirly» transcription factor is required for salicylic acid-dependent disease resistance in *Arabidopsis*. *Dev. Cell* 6: 229–240.
- Drouin, G., Daoud, H., and Xia, J. (2008). Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Mol. Phylogenet. Evol.* 49: 827–831.
- Edmondson, A.C., Song, D., Alvarez, L.A., Wall, M.K., Almond, D., McClellan, D.A., Maxwell, A., and Nielsen, B.L. (2005). Characterization of a mitochondrially targeted single-stranded DNA-binding protein in *Arabidopsis thaliana*. *Mol. Genet. Genomics* 273: 115–122.
- Emsley, P. and Cowtan, K. (2004). Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60: 2126–2132.
- Fauron, C., Allen, J., Clifton, S., and Newton, K. (2004). Plant mitochondrial genomes. In *Molecular Biology and Biotechnology of Plant Organelles*, H. Daniell and C. Chase,

- eds (Dordrecht, The Netherlands: Springer), pp. 151–177.
- Ferreira, P., Hemerly, A., de Almeida Engler, J., Bergounioux, C., Burssens, S., Van Montagu, M., Engler, G., and Inzé, D. (1994). Three discrete classes of *Arabidopsis* cyclins are expressed during different intervals of the cell cycle. *Proc. Natl. Acad. Sci. USA* 91: 11313–11317.
- Fisher, L.M., Mizuuchi, K., O’Dea, M.H., Ohmori, H., and Gellert, M. (1981). Site-specific interaction of DNA gyrase with DNA. *Proc. Natl. Acad. Sci. USA* 78: 4165–4169.
- Gellert, M., Mizuuchi, K., O’Dea, M.H., Itoh, T., and Tomizawa, J.I. (1977). Nalidixic acid resistance: A second genetic character involved in DNA gyrase activity. *Proc. Natl. Acad. Sci. USA* 74: 4772–4776.
- Gellert, M., O’Dea, M.H., Itoh, T., and Tomizawa, J. (1976). Novobiocin and coumermycin inhibit DNA supercoiling catalyzed by DNA gyrase. *Proc. Natl. Acad. Sci. USA* 73: 4474–4478.
- Gouet, P., Robert, X., and Courcelle, E. (2003). ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.* 31: 3320–3323.
- Grabowski, E., Miao, Y., Mulisch, M., and Krupinska, K. (2008). Single-stranded DNA-binding protein Whirly1 in barley leaves is located in plastids and the nucleus of the same cell. *Plant Physiol.* 147: 1800–1804.
- Hastings, P.J., Ira, G., and Lupski, J.R. (2009a). A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet.* 5: e1000327.
- Hastings, P.J., Lupski, J.R., Rosenberg, S.M., and Ira, G. (2009b). Mechanisms of change in gene copy number. *Nat. Rev. Genet.* 10: 551–564.
- Horvath, M.P. (2008). Single-stranded nucleic acid-binding proteins. In *Protein-Nucleic Acid Interactions: Structural Biology*, P.A. Rice and C.C. Correll, eds (London: Royal Society of Chemistry), pp. 91–128.
- Janska, H., Sarria, R., Woloszynska, M., Arrieta-Montiel, M., and Mackenzie, S.A. (1998). Stoichiometric shifts in the common bean mitochondrial genome leading to male sterility and spontaneous reversion to fertility. *Plant Cell* 10: 1163–1180.[Abstract/Free Full Text]

- Kanno, A., Watanabe, N., Nakamura, I., and Hirai, A. (1993). Variations in chloroplast DNA from rice (*Oryza sativa*): Differences between deletions mediated by short direct-repeat sequences within a single-species. *Theor. Appl. Genet.* 86: 579–584.
- Knoop, V. (2004). The mitochondrial DNA of land plants: Peculiarities in phylogenetic perspective. *Curr. Genet.* 46: 123–139.
- Kohl, S. and Bock, R. (2009). Transposition of a bacterial insertion sequence in chloroplasts. *Plant J.* 58: 423–436.
- Krause, K., Kilbiński, I., Mulisch, M., Rödiger, A., Schäfer, A., and Krupinska, K. (2005). DNA-binding proteins of the Whirly family in *Arabidopsis thaliana* are targeted to the organelles. *FEBS Lett.* 579: 3707–3712.
- López, E. and Blázquez, J. (2009). Effect of subinhibitory concentrations of antibiotics on intrachromosomal homologous recombination in *Escherichia coli*. *Antimicrob. Agents Chemother.* 53: 3411–3415.
- Lu, D. and Keck, J.L. (2008). Structural basis of *Escherichia coli* single-stranded DNA-binding protein stimulation of exonuclease I. *Proc. Natl. Acad. Sci. USA* 105: 9169–9174.
- Lupski, J.R. (2009). Genomic disorders ten years on. *Genome Med.* 1: 42.
- Mackenzie, S.A. (2005). The mitochondrial genome of higher plants: A target for natural adaptation. In *Plant Diversity and Evolution: Genotypic and Phenotypic Variation in Higher Plants*, R. Henry, ed (London: CABI Publishers), pp. 69–80.
- Maréchal, A. and Brisson, N. (2010). Recombination and the maintenance of plant organelle genome stability. *New Phytol.* 186: 299–317.
- Maréchal, A., Parent, J.S., Sabar, M., Véronneau-Lafortune, F., Abou-Rached, C., and Brisson, N. (2008). Overexpression of mtDNA-associated AtWhy2 compromises mitochondrial function. *BMC Plant Biol.* 8: 42.
- Maréchal, A., Parent, J.S., Véronneau-Lafortune, F., Joyeux, A., Lang, B.F., and Brisson, N. (2009). Whirly proteins maintain plastid genome stability in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* 106: 14693–14698.
- Marvo, S.L., King, S.R., and Jaskunas, S.R. (1983). Role of short regions of homology in intermolecular illegitimate recombination events. *Proc. Natl. Acad. Sci. USA* 80: 2452–2456.

- Maul, J.E., Lilly, J.W., Cui, L., dePamphilis, C.W., Miller, W., Harris, E.H., and Stern, D.B. (2002). The *Chlamydomonas reinhardtii* plastid chromosome: Islands of genes in a sea of repeats. *Plant Cell* 14 2659–2679.
- McVey, M. and Lee, S.E. (2008). MMEJ repair of double-strand breaks (director's cut): Deleted sequences and alternative endings. *Trends Genet.* 24: 529–538.
- Menke, M., Chen, I., Angelis, K.J., and Schubert, I. (2001). DNA damage and repair in *Arabidopsis thaliana* as measured by the comet assay after treatment with different classes of genotoxins. *Mutat. Res.* 493: 87–93.
- Moeykens, C.A., Mackenzie, S.A., and Shoemaker, R.C. (1995). Mitochondrial genome diversity in soybean: Repeats and rearrangements. *Plant Mol. Biol.* 29: 245–254.
- Morel, F., Renoux, M., Lachaume, P., and Alziari, S. (2008). Bleomycin-induced double-strand breaks in mitochondrial DNA of *Drosophila* cells are repaired. *Mutat. Res.* 637: 111–117.
- Newton, K.J., Gabay-Laughnan, S., and Paepe, R.D. (2004). Mitochondrial mutation in plants. In *Plant Mitochondria*, D.A. Day, A.H. Millar, and J. Whelan, eds (Dordrecht, The Netherlands: Kluwer Academic Publishers), pp. 121–141.
- Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* 302: 205–217.
- Odahara, M., Kuroiwa, H., Kuroiwa, T., and Sekine, Y. (2009). Suppression of repeat-mediated gross mitochondrial genome rearrangements by RecA in the moss *Physcomitrella patens*. *Plant Cell* 21: 1182–1194.
- Odom, O.W., Baek, K.H., Dani, R.N., and Herrin, D.L. (2008). *Chlamydomonas* chloroplasts can use short dispersed repeats and multiple pathways to repair a double-strand break in the genome. *Plant J.* 53: 842–853.
- Ogihara, Y., Terachi, T., and Sasakuma, T. (1988). Intramolecular recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species. *Proc. Natl. Acad. Sci. USA* 85: 8573–8577.
- Persky, N.S. and Lovett, S.T. (2008). Mechanisms of recombination: Lessons from *E. coli*. *Crit. Rev. Biochem. Mol. Biol.* 43: 347–370.
- Ponder, R.G., Fonville, N.C., and Rosenberg, S.M. (2005). A switch from high-fidelity to error-prone DNA double-strand break repair underlies stress-induced mutation. *Mol.*

Cell 19: 791–804.

- Prikryl, J., Watkins, K.P., Friso, G., Wijk, K.J., and Barkan, A. (2008). A member of the Whirly family is a multifunctional RNA- and DNA-binding protein that is essential for chloroplast biogenesis. *Nucleic Acids Res.* 36: 5152–5165.
- Raubeson, L.A. and Jansen, R.K. (2005). Chloroplast genomes of plants. In *Plant Diversity and Evolution: Genotypic and Phenotypic Variation in Higher Plants*, R. Henry, ed (London: CABI Publishers), pp. 45–68.
- Ravi, V., Khurana, J.P., Takagi, A.K., and Khurana, P. (2008). An update on chloroplast genomes. *Plant Syst. Evol.* 271: 101–122.
- Roa, H., Lang, J., Culligan, K.M., Keller, M., Holec, S., Cognat, V., Montane, M.H., Houlné, G., and Chabouté, M.E. (2009). Ribonucleotide reductase regulation in response to genotoxic stress in *Arabidopsis*. *Plant Physiol.* 151: 461–471.
- Rowan, B.A., Oldenburg, D.J., and Bendich, A.J. (2010). RecA maintains the integrity of chloroplast DNA molecules in *Arabidopsis*. *J. Exp. Bot.* 61: 2575–2588.
- Saha, S., Nicholson, A., and Kapler, G.M. (2001). Cloning and biochemical analysis of the tetrahymena origin binding protein TIF1: Competitive DNA binding in vitro and in vivo to critical rDNA replication determinants. *J. Biol. Chem.* 276: 45417–45426.
- Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., and Tabata, S. (1999). Complete structure of the chloroplast genome of *Arabidopsis thaliana*. *DNA Res.* 6: 283–290.
- Schumacher, M.A., Karamooz, E., Zíková, A., Trantírek, L., and Lukes, J. (2006). Crystal structures of *T. brucei* MRP1/MRP2 guide-RNA binding complex reveal RNA matchmaking mechanism. *Cell* 126: 701–711.
- Shedge, V., Arrieta-Montiel, M., Christensen, A.C., and Mackenzie, S.A. (2007). Plant mitochondrial recombination surveillance requires unusual RecA and MutS homologs. *Plant Cell* 19: 1251–1264.
- Shen, C.C., Wertelecki, W., Driggers, W.J., LeDoux, S.P., and Wilson, G.L. (1995). Repair of mitochondrial DNA damage induced by bleomycin in human cells. *Mutat. Res.* 337: 19–23.
- Shereda, R.D., Kozlov, A.G., Lohman, T.M., Cox, M.M., and Keck, J.L. (2008). SSB as an organizer/mobilizer of genome maintenance complexes. *Crit. Rev. Biochem. Mol. Biol.* 43: 289–318.

- Shereda, R.D., Reiter, N.J., Butcher, S.E., and Keck, J.L. (2009). Identification of the SSB binding site on *E. coli* RecQ reveals a conserved surface for binding SSB's C terminus. *J. Mol. Biol.* 386: 612–625.
- Small, I.D., Isaac, P.G., and Leaver, C.J. (1987). Stoichiometric differences in DNA molecules containing the *atpA* gene suggest mechanisms for the generation of mitochondrial genome diversity in maize. *EMBO J.* 6: 865–869.
- Snyder, M. and Drlica, K. (1979). DNA gyrase on the bacterial chromosome: DNA cleavage induced by oxolinic acid. *J. Mol. Biol.* 131: 287–302.
- Sugino, A., Peebles, C.L., Kreuzer, K.N., and Cozzarelli, N.R. (1977). Mechanism of action of nalidixic acid: Purification of *Escherichia coli* *nalA* gene product and its relationship to DNA gyrase and a novel nicking-closing enzyme. *Proc. Natl. Acad. Sci. USA* 74: 4767–4771.
- Tamura, K., Adachi, Y., Chiba, K., Oguchi, K., and Takahashi, H. (2002). Identification of Ku70 and Ku80 homologues in *Arabidopsis thaliana*: Evidence for a role in the repair of DNA double-strand breaks. *Plant J.* 29: 771–781.
- Unseld, M., Marienfeld, J.R., Brandt, P., and Brennicke, A. (1997). The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nat. Genet.* 15: 57–61.
- Vermel, M., Guermann, B., Delage, L., Grienemberger, J.M., Maréchal-Drouard, L., and Gualberto, J.M. (2002). A family of RRM-type RNA-binding proteins specific to plant mitochondria. *Proc. Natl. Acad. Sci. USA* 99: 5866–5871.
- Wall, M.K., Mitchenall, L.A., and Maxwell, A. (2004). *Arabidopsis thaliana* DNA gyrase is targeted to chloroplasts and mitochondria. *Proc. Natl. Acad. Sci. USA* 101: 7821–7826.
- Wang, H., Perrault, A.R., Takeda, Y., Qin, W., and Iliakis, G. (2003). Biochemical evidence for Ku-independent backup pathways of NHEJ. *Nucleic Acids Res.* 31: 5377–5388.
- Weigel, D. and Glazebrook, J. (2002). *Arabidopsis: A Laboratory Manual*. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).
- Woelfle, M.A., Thompson, R.J., and Mosig, G. (1993). Roles of novobiocin-sensitive topoisomerases in chloroplast DNA replication in *Chlamydomonas reinhardtii*. *Nucleic Acids Res.* 21: 4231–4238.

- Wolfe, K.H., Li, W.H., and Sharp, P.M. (1987). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci. USA* 84: 9054–9058.
- Woloszynska, M. (2010). Heteroplasmy and stoichiometric complexity of plant mitochondrial genomes—Though this be madness, yet there's method in't. *J. Exp. Bot.* 61: 657–671.
- Xiong, J.Y., Lai, C.X., Qu, Z., Yang, X.Y., Qin, X.H., and Liu, G.Q. (2009). Recruitment of AtWHY1 and AtWHY3 by a distal element upstream of the kinesin gene AtKP1 to mediate transcriptional repression. *Plant Mol. Biol.* 71: 437–449.
- Yakisich, J.S., Sandoval, P.Y., Morrison, T.L., and Kapler, G.M. (2006). TIF1 activates the intra-S-phase checkpoint response in the diploid micronucleus and amitotic polyploid macronucleus of *Tetrahymena*. *Mol. Biol. Cell* 17: 5185–5197.
- Ye, J. and Sayre, R.T. (1990). Reduction of chloroplast DNA content in *Solanum nigrum* suspension cells by treatment with chloroplast DNA synthesis inhibitors. *Plant Physiol.* 94: 1477–1483.
- Yoo, H.H., Kwon, C., Lee, M.M., and Chung, I.K. (2007). Single-stranded DNA binding factor AtWHY1 modulates telomere length homeostasis in *Arabidopsis*. *Plant J.* 49: 442–451.
- Zaegel, V., Guermann, B., Le Ret, M., Andrés, C., Meyer, D., Erhardt, M., Canaday, J., Gualberto, J.M., and Imbault, P. (2006). The plant-specific ssDNA binding protein OSB1 is involved in the stoichiometric transmission of mitochondrial DNA in *Arabidopsis*. *Plant Cell* 18: 3548–3563.

Table 3-I. Representative DNA rearrangements observed in *Arabidopsis* organellar genomes

Event	Plant Treatment ^a	Plant Genotype	Junction Endpoints ^b	Junction Sequences ^c	Length of Junctions (Mismatches)	Junction Types ^d
NHEJ-like DNA junction						
M3-2	CIP 0.75 μ M	Col-0	55001 Recombinant 65518	cgaagagagc <u>cataccggaa</u> tgagcgcgtg <u>cataccggaa</u> tgagcgcgtg tactgtggca	0 bp	NHEJ-like
MHMR DNA junction						
M3-3	CIP 0.75 μ M	KO2	53421 Recombinant 65476	gctccgtcca <u>TGGAGCGTATG</u> attttccttg ctctaaagct <u>TGGAGCGTATG</u> attttccttg ctctaaagct <u>TGGAGCGTATG</u> gtaacggtta	(0) 11 bp (0)	MHMR
MHMR DNA junction containing a mismatch in one parental strand						
M6-1	CIP 0.25 μ M	KO2	234779 Recombinant 246725	gatgaggatg <u>AAGTAAAGGAATT</u> actctggttg acgaaagtga <u>AAGTAAAGGAATT</u> actctggttg acgaaagtga <u>AAGTAAAGGAATT</u> gtcctggccct	(0) 13 bp (1)	MHMR*
MHMR DNA junction containing a mismatch in both parental strands						
M7-7	CIP 0.75 μ M	Col-0	304900 Recombinant 316104	tggaatgctc <u>TCGTTGTTAAAAA</u> ctacgagcgg caggggggat <u>TCTTGTCGATAAAA</u> ctacgagcgg caggggggat <u>TCTTGTCGATAAAA</u> ggcattogcgg	(4) 15 bp (1)	MHMR*
Same MHMR DNA junction containing different mismatches						
M3-6	CIP 0.75 μ M	KO2	54234 Recombinant 65096	taggaatctt <u>AGGTTACTCCAG</u> ccttctgaag ccactaggta <u>AGGTTACTCCAG</u> ccttctgaag ccactaggta <u>AGGTTACTCCAG</u> agcggaggtat	(0) 13 bp (1)	MHMR*
M3-7	CIP 0.75 μ M	KO2	54234 Recombinant 65096	taggaatctt <u>AGGTTACTCCAG</u> ccttctgaag ccactaggta <u>AGGTTACTCCAG</u> ccttctgaag ccactaggta <u>AGGTTACTCCAG</u> agcggaggtat	(1) 13 bp (0)	MHMR*
MHMR DNA junctions with homopolymeric tracts						
M12-1	CIP 0.25 μ M	Col-0	29609 Recombinant 80994	tttttccctt <u>CCCCCCCCCA</u> aaaaaaaa attagacgtc <u>CCCCCCCCCA</u> aaaaaaaaag attagacgtc <u>CCCCCCCCCA</u> tgctacggtt	(0) 10 bp (0)	MHMR
P10-14	CIP 0.75 μ M	KO1/3	59719 Recombinant 50843	tactttgaaa <u>ACAAAAAAA</u> ggaaagtcat gagtggggat <u>ACAAAAAAA</u> ggaaagtcat gagtggggat <u>ACAAAAAAA</u> aaagaaagta	(0) 10 bp (0)	MHMR
Molecules containing multiple MHMR						
P6-3a	CIP 0 μ M	KO1/3	19786 Recombinant 110850	catatctcaa <u>TAGATTGGCGT-AATTTT</u> ctaccgcatg caaaattttg <u>TAGATTGGCGT-AATTTT</u> ctaccgcatg caaaattttg <u>TAGATTGGTATGAATTTT</u> gagaaatgca	(0) 18 bp (3)	MHMR*
P6-3b	CIP 0 μ M	KO1/3	19867 Recombinant 122003	taattatggg <u>AGAACTTGATCGTTAAATGAT</u> aatttatgca gccagccctt <u>AGAACTTGATCGTTAAATGAT</u> aatttatgca gccagccctt <u>AGAACTT-ATCGTTAAAA-GAT</u> catcaattcc	(0) 22 bp (3)	MHMR*

A complete list of DNA rearrangements can be found in Supplemental Table 1 online.

^a Treatment of plants with ciprofloxacin (CIP) or novobiocin (NOV) at the indicated concentrations.

^b The junction endpoint is the position of the first nucleotide preceding the DNA junction. The accession numbers for the mitochondrial and plastid genomes are NC_001284 (Unseid et al., 1997) and NC_000932 (Sato et al., 1999), respectively.

^c Alignment of the sequenced DNA junction (center) with homologous sequences (top and bottom). Mismatches are underlined.

^d Product types were classified according to junction length with NHEJ-like for junction of less than five nucleotides, MHMR for junctions of five or more nucleotides and MHMR* for junctions of five or more nucleotides containing at least one mismatch.

Table 3-II. Data collection and refinement statistics

Protein	WHY2 Free Form	WHY2 -ERE ₃₂	WHY2 -dT ₄₂	WHY2 -cERE ₃₂	WHY2 -rcERE ₃₂
Beamline/Wavelength	NSLS- X25/1.08 Å	NSLS- X29/1.08 Å	NSLS- X25/1.08 Å	APS-22ID/1.10 Å	APS-22ID/1.10 Å
Space group	F432	F432	F432	F432	F432
Cell dimensions :					
<i>a</i> (Å)	164.58	167.16	166.51	166.93	166.67
Resolution (Å)	50–2.20 (2.28–2.20)	50–2.20 (2.28–2.20)	50–2.65 (2.74–2.65)	50–2.70 (2.80–2.70)	50–2.35 (2.43–2.35)
Total Reflections	380467	382748	118110	178903	253462
Unique Reflections	10208	10665	6188	5865	8716
R _{sym} (%)	7.6 (92)	5.1 (56)	11.8 (87)	9.0 (68)	8.8 (75)
I / σ	16.9 (3.8)	23.1 (3.8)	9.1 (3.0)	12.7 (2.2)	16.5 (2.3)
Completeness (%)	100.0 (99.9)	99.8 (98.7)	100.0 (100.0)	99.6 (96.0)	99.5 (95.4)
Multiplicity	37.3 (19.5)	35.9 (9.9)	19.1 (12.7)	30.5 (5.9)	29.1 (6.5)
Refinement Statistics					
Resolution (Å)	50–2.20	50–2.20	50–2.65	50–2.70	50–2.35
Reflections (total/test) ^a	9873/532	9843/752	6012/603	5625/628	8437/900
R _{work} / R _{free} (%)	20.32/25.08	22.67/26.18	21.15/25.79	20.75/25.71	21.70/26.26
Number of atoms					
Protein	1267	1267	1267	1267	1267
DNA/Phosphate	10	180	180	189	189
Water	69	36	60	56	60
B-factors					
Protein	50.9	52.5	53.5	60.7	61.8
DNA/Phosphate	56.8	62.9	61.2	72.6	75.0
Water	53.5	46.1	52.8	58.4	62.0
RMSDs					
Bond lengths (Å)	0.013	0.008	0.006	0.006	0.010
Bond angles (°)	1.414	1.099	0.997	0.953	1.321
Ramachandran^b					
Favored (%)	96.8	97.5	95.0	97.5	95.6
Outliers (%)	0	0	0	0	0

Values in parentheses are for highest-resolution shell.

$R_{\text{sym}} = \sum_{hkl} \sum_i |I_{hkl,i} - \langle I_{hkl} \rangle| / \sum_{hkl,i} \langle I_{hkl,i} \rangle$, where $I_{hkl,i}$ is the intensity of an individual measurement of the reflection with Miller indices hkl and $\langle I_{hkl} \rangle$ is the mean intensity of that reflection.

$R_{\text{work}} = \sum_{hkl} |I_{\text{F}_o} - I_{\text{F}_c}| / \sum_{hkl} I_{\text{F}_o}$, where I_{F_o} is the observed structure-factor amplitude and I_{F_c} is the calculated structure-factor amplitude.

R_{free} is the R factor based on at least 500 test reflections that were excluded from the refinement.

^a Reflections with $F_o > 0$

^b MolProbity analysis (Davis et al., 2007)

Table 3-III. Analysis of the binding of St-WHY2 to four different DNA sequences by EMSA.

DNA name	DNA sequence	$K_{d,app}$ (nM)	n^a
ERE ₃₂	TGTCATTTTTGTCATTTTTGTCATTTTTGTCA	3.35 ± 0.95	5
dT ₃₂	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	4.45 ± 0.87	5
rcERE ₃₂	ACAGTAAAAACAGTAAAAACAGTAAAAACAGT	4.29 ± 0.68	3
cERE ₃₂	TGACAAAAATGACAAAAATGACAAAAATGACA	5.73 ± 1.33	5

^a: number of independent measurements (each independent measurement was done in duplicates)

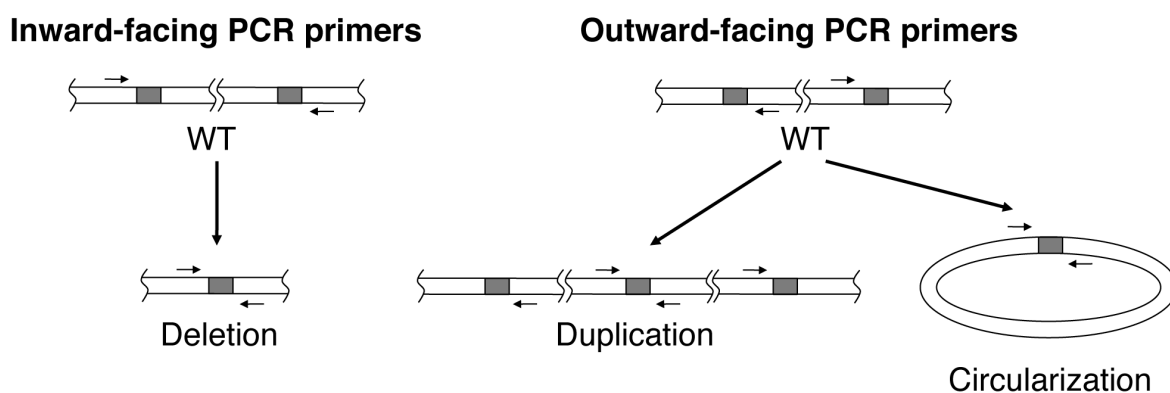


Figure 3-1. Outward-facing PCR primers are used to monitor DNA duplication/circularization events, whereas inward-facing PCR primers detect deletions. In both cases, PCR amplification occurs only if a DNA rearrangement brings together the annealing sites of the primers. Gray areas represent repeated sequences.

Figure 3-2. DNA rearrangements accumulate in plastids following treatment with the gyrase inhibitor ciprofloxacin.

(A) and (E) Phenotypic effects of various concentrations of ciprofloxacin (A) or novobiocin (E) on wild-type (WT) and *why1 why3 Arabidopsis* plants. Plants were grown for 3 weeks on solid media containing the indicated concentrations of ciprofloxacin (CIP) or novobiocin (NOV).

(B) and (F) Histograms showing the average \pm SD of plants with etiolated/variegated leaves at each ciprofloxacin (B) or novobiocin (F) concentration for wild-type and *why1 why3* genotypes. At least three independent experiments were done. Plants with partially or fully white first true leaves were scored as etiolated/variegated. ND, etiolated/variegated leaves could not be counted at 0.75 μ M ciprofloxacin and 100 μ M novobiocin because they were very small or absent.

(C) and (G) Electrophoretic analysis of representative PCR performed with 13 outward- or inward-facing plastid genome-directed PCR primers on total leaf DNA of wild-type and *why1 why3* plants treated with ciprofloxacin (C) or with 10 outward- or inward-facing plastid genome-directed PCR primers on total leaf DNA of wild-type and *why1 why3* plants treated with novobiocin (G). Low cycle amplification of the *YCF2* plastid gene was used as a loading control. The oligonucleotides used for each PCR are indicated. Individual bands were cut from the gel, cloned, and sequenced. DNA rearrangements are listed Table 3-SI.

(D) and (H) Histograms showing the number of PCR products in wild-type and *why1 why3* plants as a function of ciprofloxacin (D) or novobiocin (H) concentration.

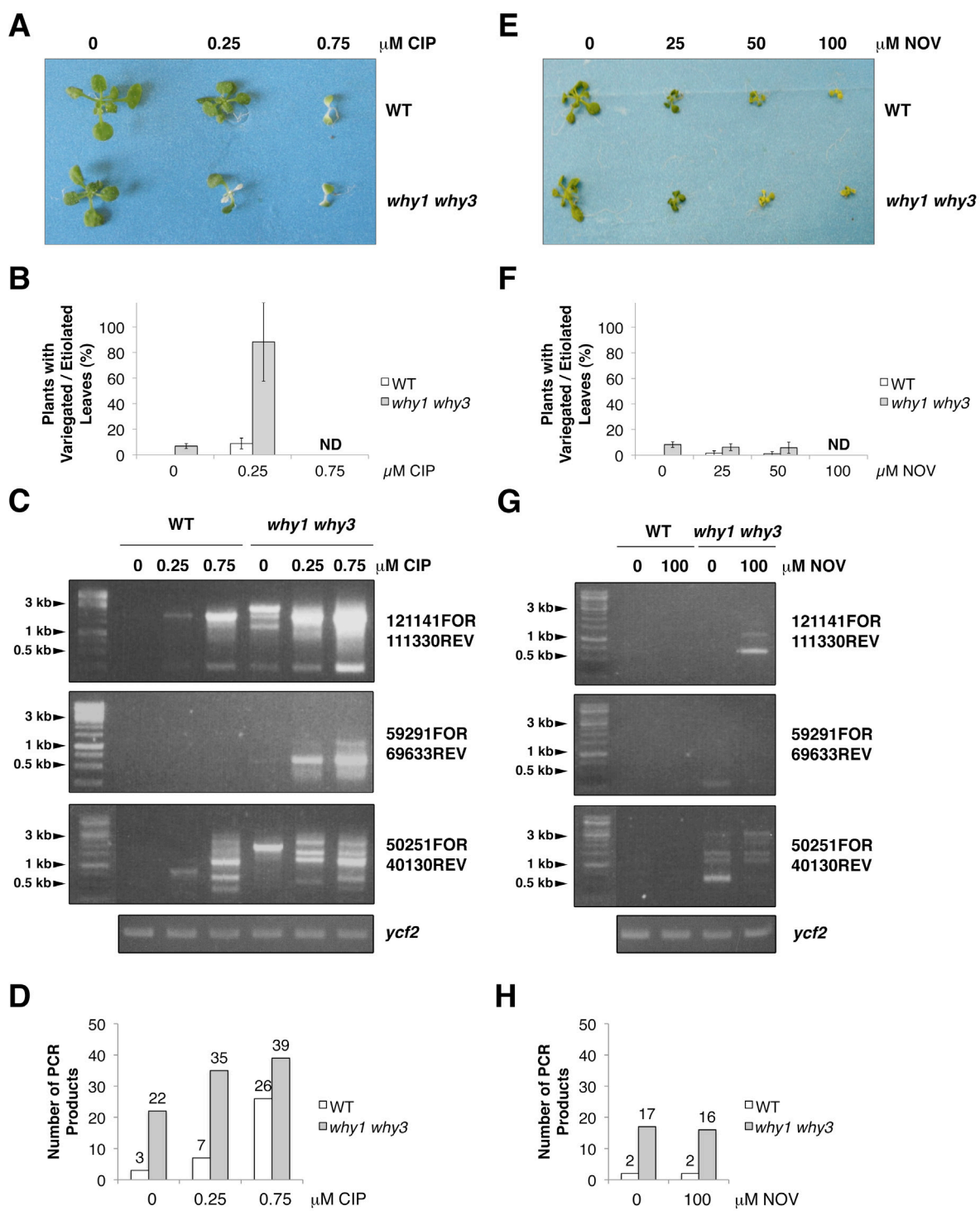


Figure 3-3. DNA rearrangements accumulate in mitochondria following treatment with the gyrase inhibitor ciprofloxacin.

(A) and (E) Phenotypic effects of various concentrations of ciprofloxacin (A) or novobiocin (E) on wild-type (WT) and *why2-1 Arabidopsis* plants. Plants were grown for 3 weeks on solid media containing the indicated concentrations of ciprofloxacin (CIP) or novobiocin (NOV).

(B) and (F) Histograms showing the average \pm SD of plants with etiolated/variegated leaves at each ciprofloxacin (B) or novobiocin (F) concentration for wild-type and *why2-1* genotypes. At least three independent experiments were done. Plants with partially or fully white first true leaves were scored as etiolated/variegated. ND, etiolated/variegated leaves could not be counted at 0.75 μ M ciprofloxacin and 100 μ M novobiocin because they were very small or absent.

(C) and (G) Electrophoretic analysis of representative PCR performed with 20 outward- or inward-facing mitochondrial genome-directed PCR primers on total leaf DNA of wild-type and *why2-1* plants treated with ciprofloxacin (C) or with 10 outward- or inward-facing mitochondrial genome-directed PCR primers on total leaf DNA of wild-type and *why2-1* plants treated with novobiocin (G). Low cycle amplification of the *COX1* mitochondrial gene was used as a loading control. The oligonucleotides used for each PCR are indicated. Individual bands were cut from the gel, cloned, and sequenced. DNA rearrangements are listed in Table 3-SI.

(D) and (H) Histograms showing the number of PCR products in wild-type and *why2-1* plants as a function of ciprofloxacin (D) or novobiocin (H) concentration.

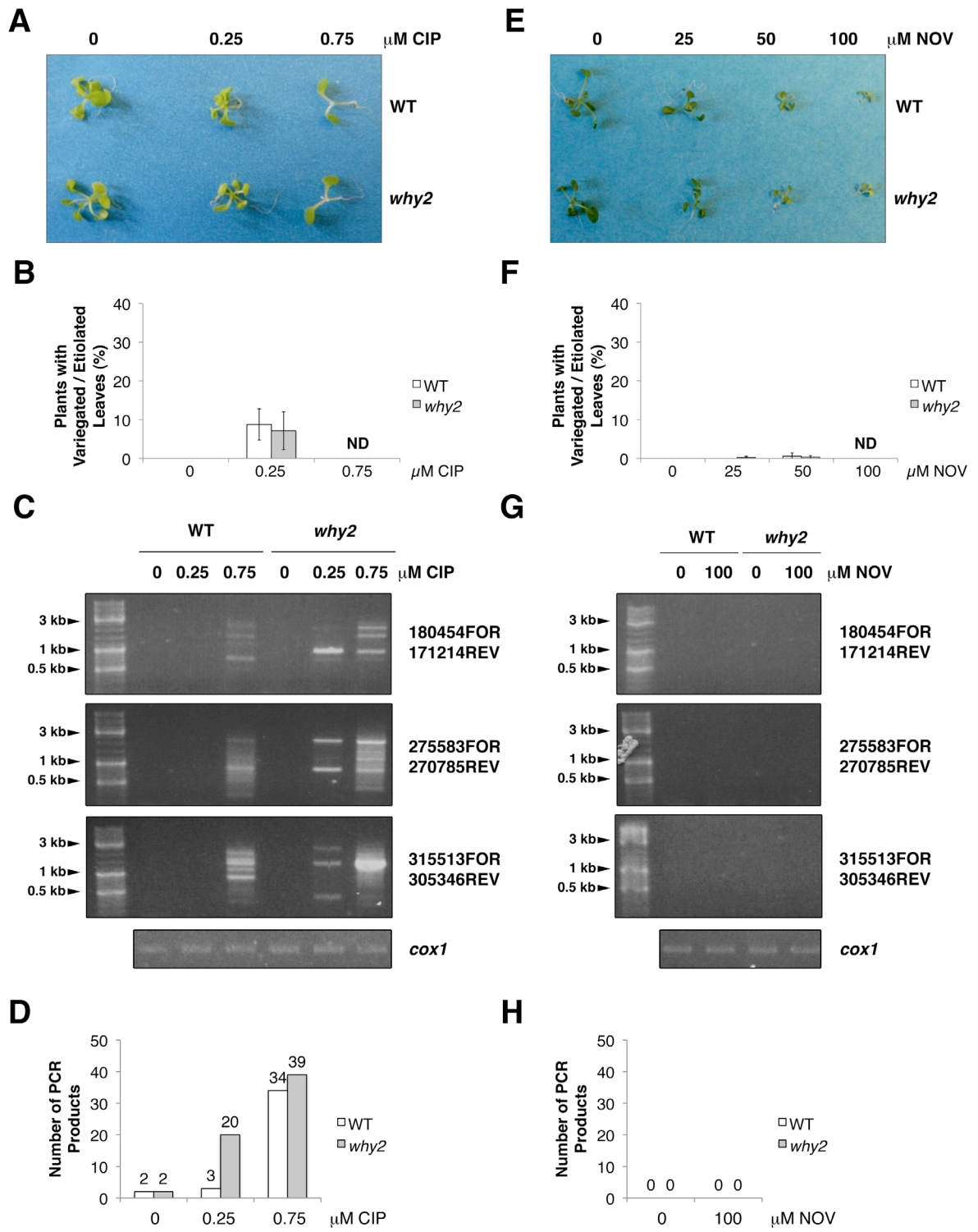


Figure 3-4. Crystal structures of St-WHY2 in the free form and bound to ERE₃₂ at 2.2-Å resolution.

(A) Schematic representation of St-WHY2. Filled boxes indicate the position of the mitochondria transit peptide (mTP), the Whirly domain, and the acidic/aromatic C-terminal tail (CT). A green dotted line represents the construct used for structure determination.

(B) Overall view of St-WHY2 in the free form in cartoon representation. The tetramer was generated by applying crystallographic fourfold symmetry. Protomers are colored in yellow, orange, pink, and green.

(C) Surface representation of the protein moiety in the St-WHY2-ERE₃₂ complex. The tetramer was generated by applying the crystallographic symmetry along the fourfold axis. Difference electron density was calculated from a $F_o - F_c$ simulated annealing omit map encompassing DNA, contoured at 3.0σ , and colored in green. The density was carved at 8 Å around the protein model.

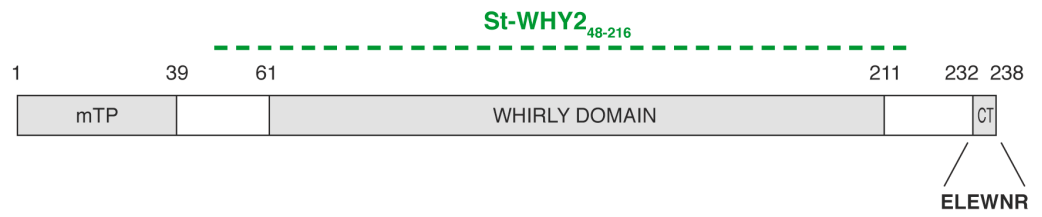
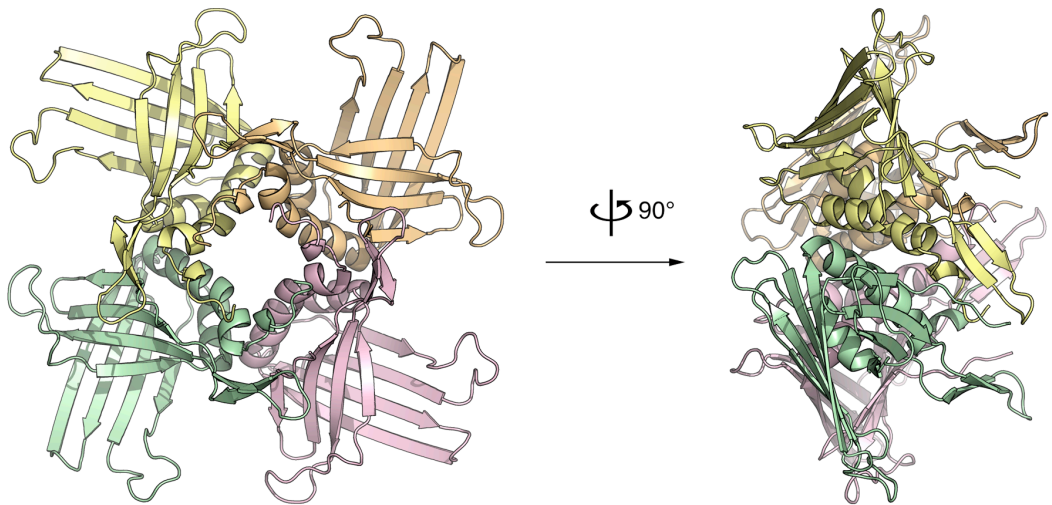
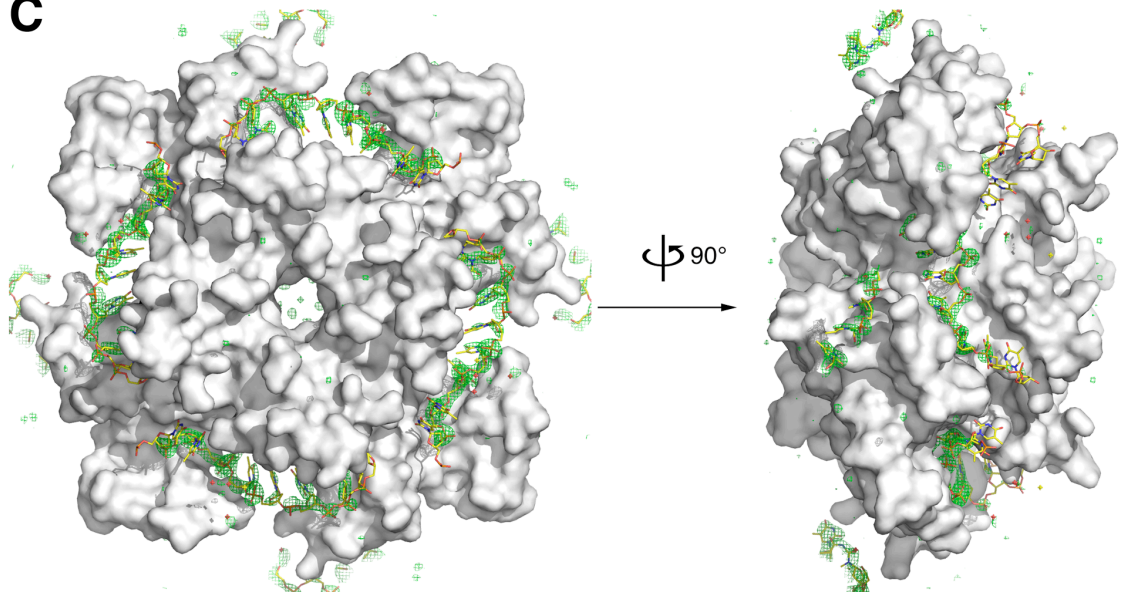
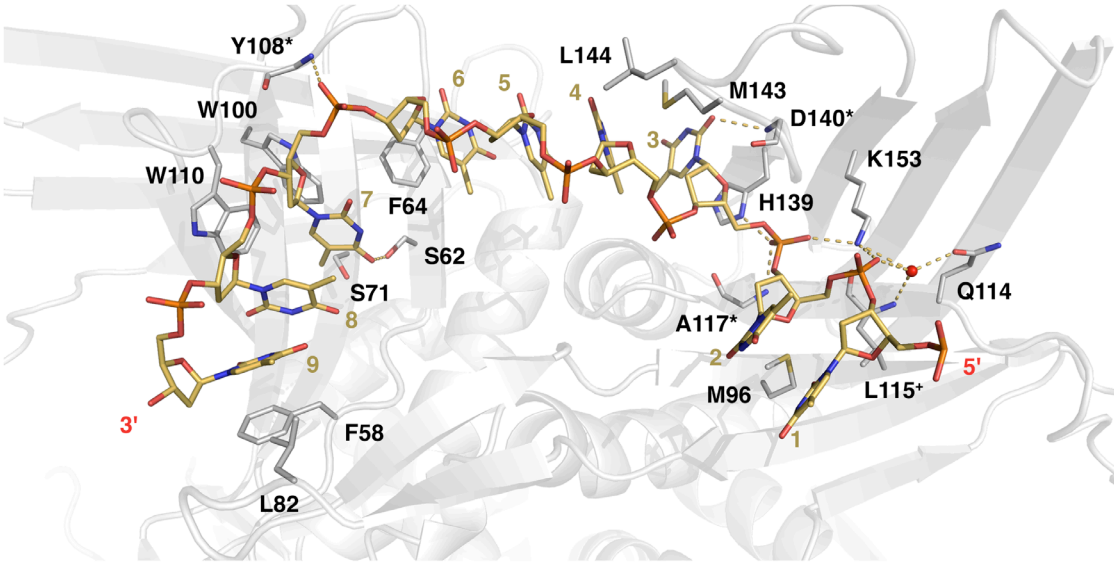
A**B****C**

Figure 3-5. Mechanism of ssDNA binding of Whirly proteins.

(A) Protein–DNA interactions in the St-WHY2-ERE₃₂ complex. The DNA and the DNA-interacting residues are in stick representation with carbon atoms colored in yellow and in gray, respectively. Nucleotides and protein residues are labeled. Asterisks, residues that contact DNA through their main-chain. +, residues that contact the DNA through both its main chain and side chain. Hydrogen bonds are represented as yellow dashed lines. A water molecule is represented as a red sphere.

(B) Sequence alignment of the Whirly domain of Whirly proteins from *Arabidopsis* (At-WHY1, At-WHY2, and At-WHY3) and *S. tuberosum* (St-WHY1 and St-WHY2) with schematic secondary structural elements from St-WHY2. Secondary structure conformations are denoted at the top of the sequence alignment. α , α -helix; β , β -strand; η , 3_{10} helix. Similarity above 70% is depicted in yellow, whereas perfect conservation is depicted in red. Blue stars underneath the sequence alignment indicate residues of St-WHY2 that interact with ssDNA. The alignments were made using T-coffee (Notredame et al., 2000) and the figure prepared using ESPript (Gouet et al., 2003).

A



B

<i>StWHY2</i>				β1		β2		β3		β4		β5		η1	
<i>StWHY2</i>	55	GRV	FAPY	SVFKGKAAL	SAEPRLPTFNR	LD	SGGV	KLNR	RGVIMLT	FW	PSVGERKYDWEKRQ				
<i>AtWHY2</i>	52	GRL	FAPY	SIFKGKAAL	SVLPVPSFTE	ID	SGNLR	RIDRR	GS	LMMTF	MPAIGERKYDWEKKQ				
<i>AtWHY1</i>	79	ARF	YVGH	SIYKGKAAL	TVDPRAPEFVA	LD	SGAF	KL	SKDGF	LLLOQ	FAPSA	GVROQYDWSKKQ			
<i>AtWHY3</i>	83	PRF	YVGH	SIYKGKAAL	TIEPRAPEFVA	LE	SGAF	KL	TKEGF	LLLOQ	FAPAA	GVROQYDWSRKQ			
<i>StWHY1</i>	90	PKV	FVGY	SIYKGKAAL	TVEPRSP	EFSP	LD	SGAF	KL	SREG	VMLOQ	FAPAA	GVROQYDWSRKQ		
				★	★	★		★			★	★		★	★

<i>StWHY2</i>			β6		α1		β7		β8		β9				
<i>StWHY2</i>	115	LF	FALS	TEV	GSLISM	GTRD	S	EFFHDP	SML	SS	NAGQVRKSL	SIKPN	AD	SGSYFI	SLSVNN
<i>AtWHY2</i>	112	KF	FALSP	TEV	GSLISM	GSKD	S	EFFHDP	SMK	SS	NAGQVRKSL	SVKPHA	D	SGSYFI	SLSVNN
<i>AtWHY1</i>	139	VF	SLSV	TEI	GTLSI	GPRES	S	EFFHDP	FKG	KS	DEGKVRKVL	KVEPL	D	SGSHFF	NLSVQN
<i>AtWHY3</i>	143	VF	SLSV	TEI	GNLVS	GPRES	S	EFFHDP	FKG	KG	DEGKVRKVL	KVEPL	D	SGSRFF	NLSVQN
<i>StWHY1</i>	150	VF	SLSV	TEI	GSII	SLGAKD	S	EFFHDP	NKG	RS	DEGRVRKVL	KVEPL	D	SGSHFF	NLSVQN
			★	★				★	★	★		★			

<i>StWHY2</i>				β10		α2		η2							
<i>StWHY2</i>	175	NNL	K	TNDRFT	VPVT	AEFAVMRTAFS	FALPH	IMGWDRFT	NR						
<i>AtWHY2</i>	172	SIL	K	TNDYFV	VPVT	KAFAVMKTAFA	SFALPH	IMGWNRLT	GH						
<i>AtWHY1</i>	199	KLV	N	VDESIY	IPIT	RAEFVAVLISAFNFVLP	YLIGW	HAFANS							
<i>AtWHY3</i>	203	KLL	N	VDESYY	IPIT	KAFAVAVLISAFNFVLP	PHLIGW	SAFANS							
<i>StWHY1</i>	210	KLI	N	LDENIY	IPVT	KAFAVAVLVSAFNFVMP	YLLGW	H	TAVNS						

Figure 3-6. WHY2 binds ssDNA with limited sequence specificity.

(A) Representative EMSA results showing the binding of St-WHY2 to four different ssDNA sequences. Increasing amounts of WHY2 were incubated with target oligonucleotides ERE₃₂, dT₃₂, rcERE₃₂, or cERE₃₂ and the complexes resolved on a 10% (w/v) polyacrylamide gel. The sequences of these oligonucleotides can be found in Table 3-III.

(B) Crystal structures of four different ssDNA sequences bound by St-WHY2. DNA molecules are presented as stick models with carbon atoms colored in yellow. F_o-F_c simulated annealing omit maps encompassing the entire DNA are contoured at 2.5 σ (colored in gray) or at 5 σ (colored in green). DNA molecules are presented in the same order as in (A).

(C) and (D) Interactions between St-WHY2 and the edges of T3 (left panel) and A3 (right panel) (C) or the edges of T7 (left panel) and A7 (right panel) (D) in the WHY2-ERE₃₂ and WHY2-rcERE₃₂ structures, respectively. The representation and the orientation of the molecule are similar, thus revealing that compensating interactions enable WHY2 to bind DNA nucleobases that differ in size and in functional groups at these positions. A red sphere corresponds to a water molecule.

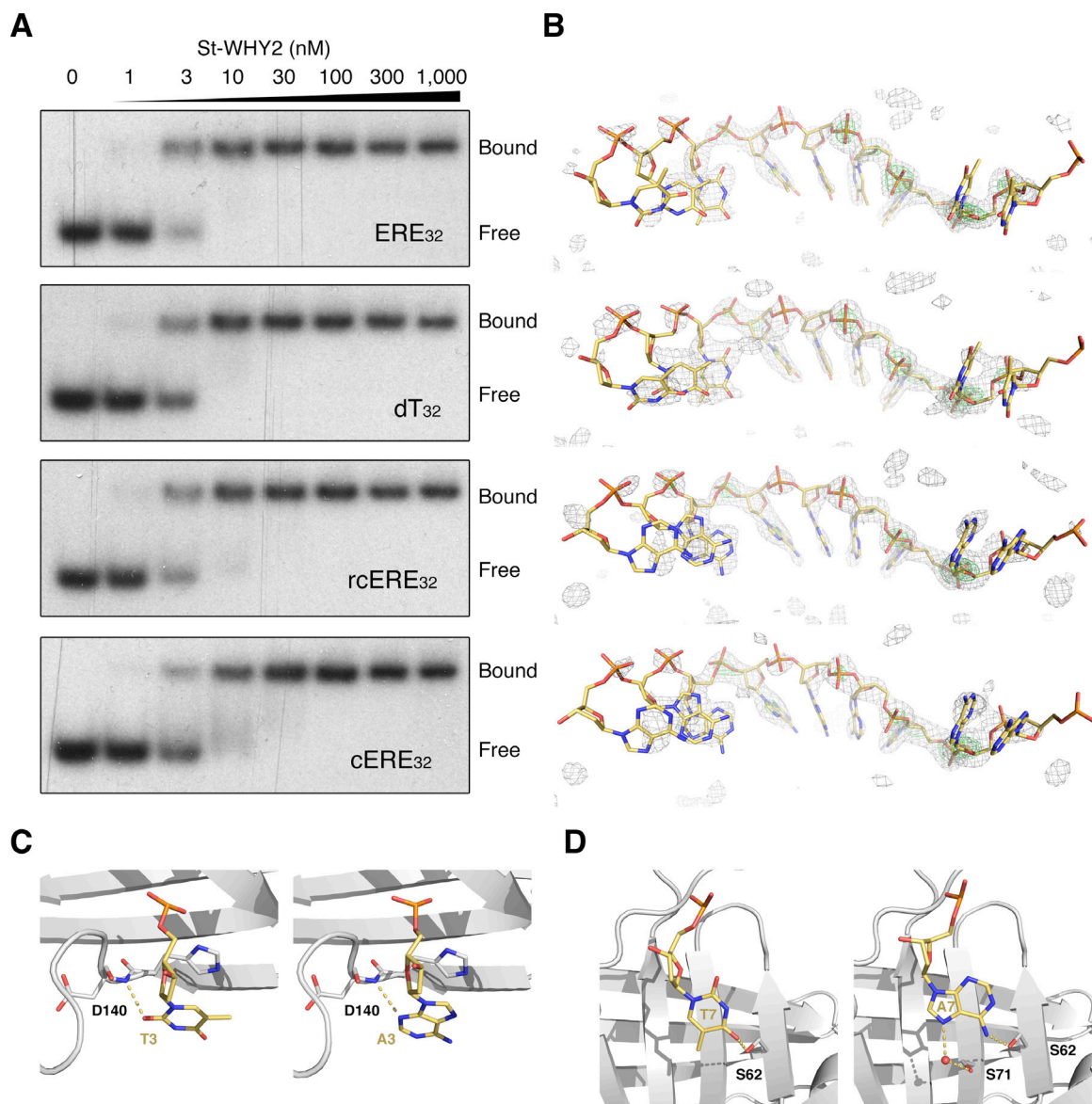


Figure 3-7. WHY2 binds single-stranded overhangs, destabilizes dsDNA, and protects ssDNA against nuclease degradation.

(A) Representative EMSA results showing the binding of St-WHY2 to DNA duplexes with or without single-stranded dT₈ or dT₁₆ overhangs. Fifty nanomolar of WHY2 were incubated with target radiolabeled oligonucleotides and the complexes resolved on a 10% (w/v) polyacrylamide gel. Diagrams at the bottom of the gels illustrate the DNA used in the assay. An asterisk indicates the strand that is radiolabeled.

(B) WHY2 destabilizes a DNA duplex. A DNA duplex with one strand radiolabeled was incubated with increasing amounts of St-WHY2 or BSA. After protein denaturation, DNA was resolved on a 7.5% (w/v) polyacrylamide gel.

(C) WHY2 protects DNA against mung bean nuclease degradation. Phage M13mp18 ssDNA either alone or prebound with St-WHY2 at a 1:10 protein/nucleotide ratio was incubated with mung bean nuclease for the indicated amount of time. After protein denaturation, DNA was resolved on an agarose gel. Black/white inverted images are shown. M represents the molecular weight markers.

(D) WHY2 protects DNA against the exonuclease activity of T4 DNA Polymerase. Radiolabeled ssDNA or dsDNA with a 16-nucleotide 3'-overhang was complexed with St-WHY2 and then incubated with T4 DNA polymerase for the indicated amount of time. After protein denaturation, DNA was resolved on a 10% (w/v) polyacrylamide gel. ssDNA, YG16_3; 3'-dsDNA, YG16_3-YC duplex; dsDNA, YG-YC duplex.

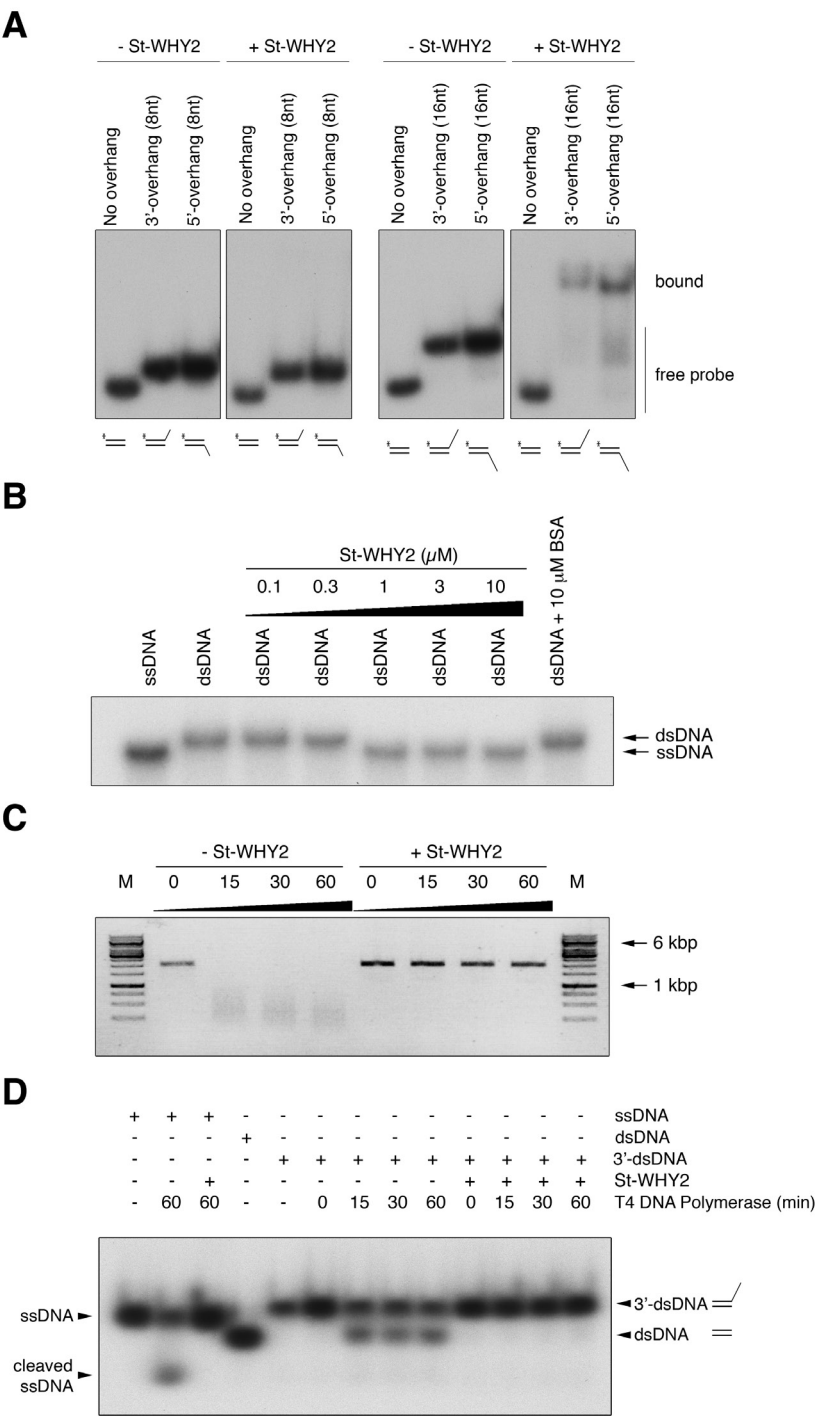
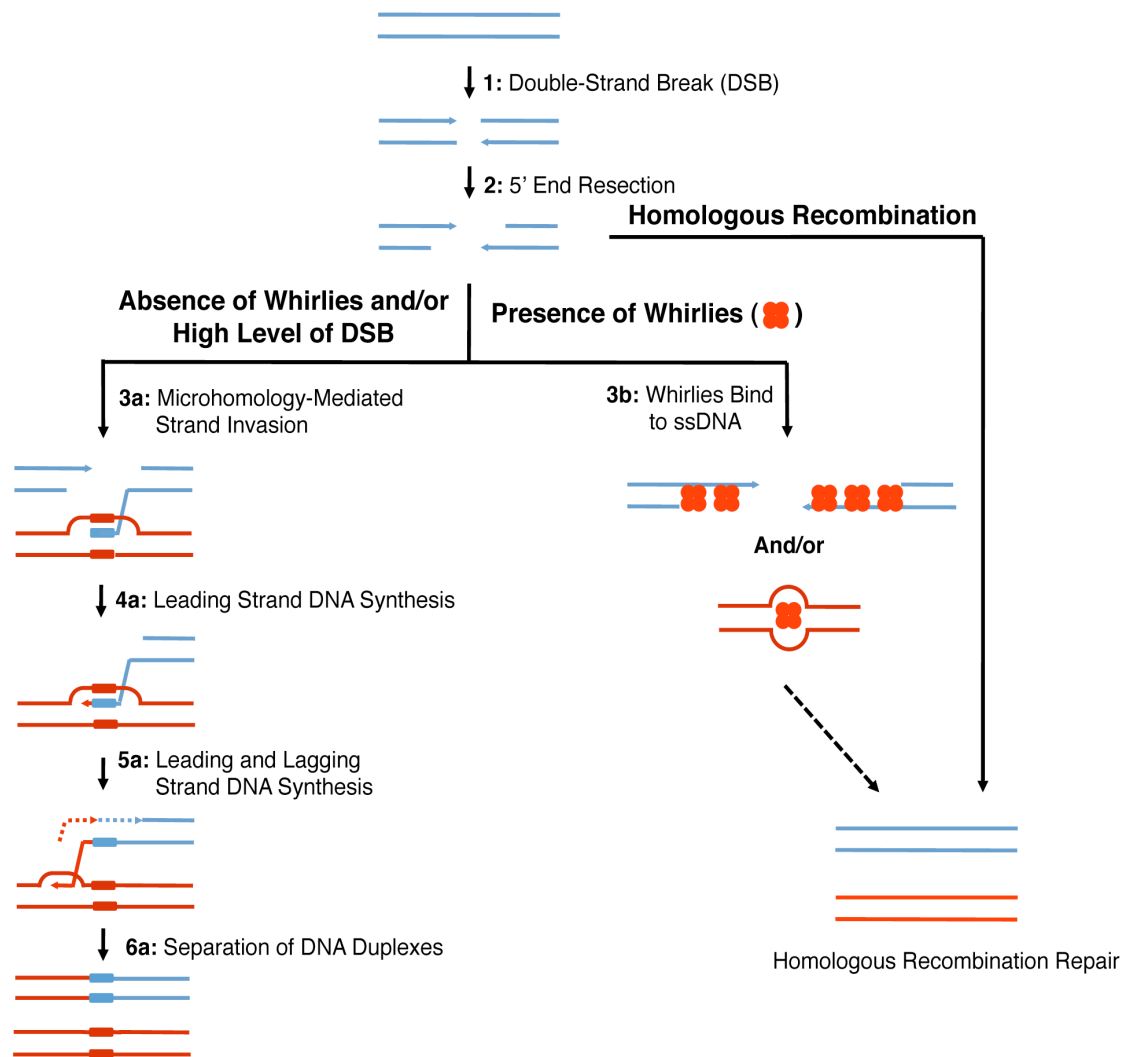


Figure 3-8. Model for the repair of organellar double-strand breaks in the absence or presence of Whirly proteins.

Upon formation of a DSB (1), the 5' end of the broken DNA molecule is resected from the break, exposing a 3' tail (2). At this step, the break can be repaired through homologous recombination in a Whirly-independent manner. Alternatively, if the Whirlies are absent or the repair machinery is overloaded due to numerous DSBs, the 3' tail can anneal to any exposed ssDNA through microhomologies (3a). A D-loop forms and DNA polymerization proceeds from the microhomology junction (4a). A replication fork is established and lagging strand synthesis initiates while leading strand synthesis continues (5a). DNA synthesis continues until the end of the chromosome is reached (6a). Alternatively, if the Whirlies are present and the DSB level is low (3b), Whirlies could bind and protect ssDNA (either the 3' tail and/or any exposed ssDNA), thereby promoting homologous recombination and accurate DNA repair. Arrowheads represent 3' ends; a box symbolizes the microhomology between broken and unbroken DNA molecules; dashed arrows in tandem represent lagging strand synthesis.



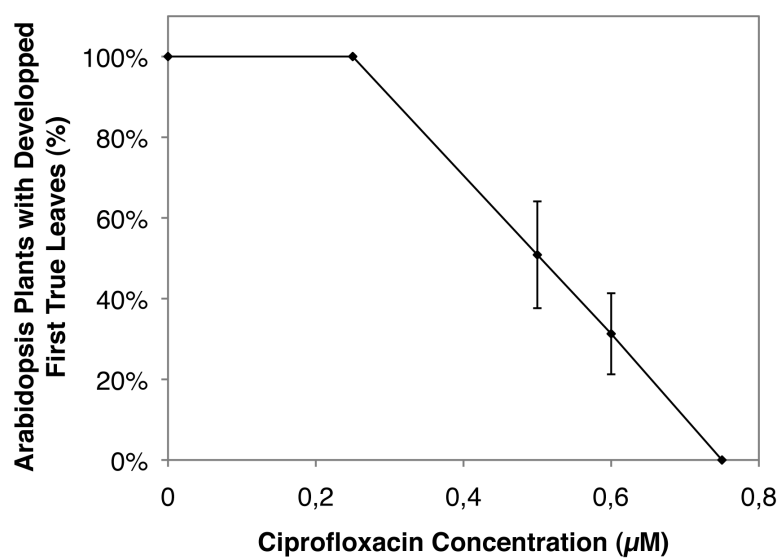


Figure 3-S1. Effect of ciprofloxacin treatment on *Arabidopsis* plant survival.

Graph showing the average percentage \pm SD of WT *Arabidopsis* plants that developed first true leaves as a function of ciprofloxacin concentration. A LD_{50} of 0.5 μ M ciprofloxacin was thus obtained for WT plants. More than 100 plants were examined for each ciprofloxacin concentration.

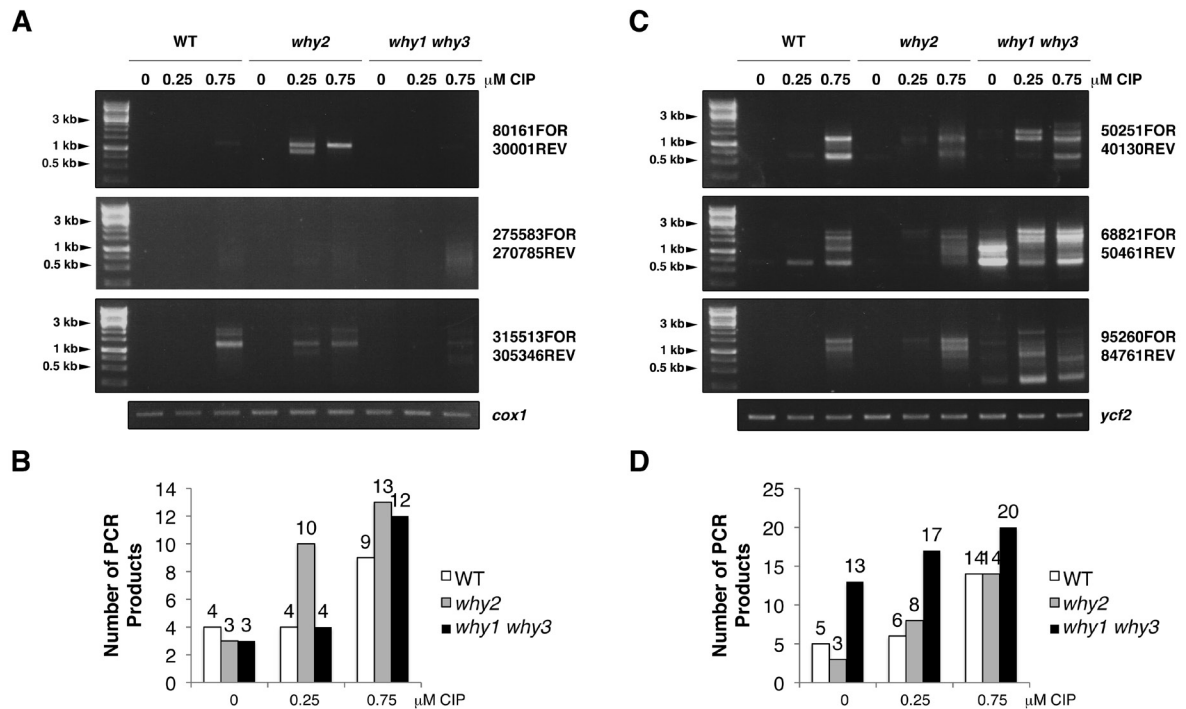


Figure 3-S2. DNA rearrangements accumulate primarily in plastids of plants lacking plastid-targeted Whirlies and in the mitochondria of plants lacking mitochondria-targeted Whirlies.

(A) and (C) Electrophoretic analysis of representative PCR performed with 8 outward- or inward-facing (A) mitochondria genome-directed and (C) plastid genome-directed PCR primers on total leaf DNA of WT, *why2-1* and *why1 why3* plants treated with ciprofloxacin. Low cycle amplifications of the *cox1* mitochondrial gene or the *ycf2* plastid gene were used as loading controls. The oligonucleotides used for each PCR are indicated.

(B) and (D) Histograms showing the number of PCR products in the mitochondrial (B) and plastid (D) genome of WT, *why2-1* and *why1 why3* plants as a function of ciprofloxacin concentration.

Figure 3-S3. DNA gel blot detection of DNA rearrangements in plastids following plant treatment with ciprofloxacin.

(A) Upper panel: Schematic representation of the microhomology-mediated DNA rearrangement occurring between two loci of the chloroplast genome of *Arabidopsis* (accession number NC_000932). The DNA rearrangement was isolated by PCR using the primers 4434REV and 15643FOR. The PCR product was used as a probe for the DNA gel blot presented in the middle and lower panels and is represented by a thick gray line. The HindIII fragments corresponding to the bands in the DNA gel blot in the middle panel are delimited by a “H”. The black boxes designate the microhomologous DNA sequences that mediated the rearrangement.

Middle panel: DNA gel blots of total DNA from WT and *why1 why3* plants treated with ciprofloxacin. The total DNA was digested with HindIII and hybridized with the probe described in the upper panel. The predicted sizes of the bands corresponding to the WT and rearranged DNA molecules are indicated. R designates the rearranged DNA molecules.

Lower panel: Detection of subgenome-length molecules in DNA gel blots of total undigested DNA from WT and *why1 why3* plants treated with ciprofloxacin using the same probe described in the upper panel. Asterisks indicate the position of rearranged DNA molecules.

(B) Same as in (A) but using as a probe a PCR product amplified with primers 49386REV and 68817FOR.

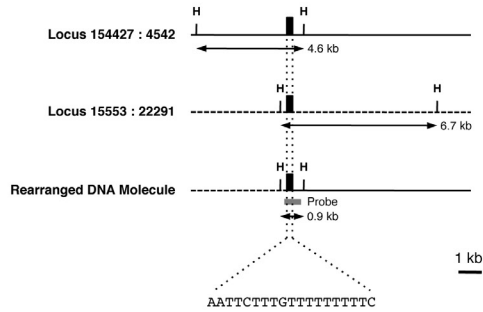
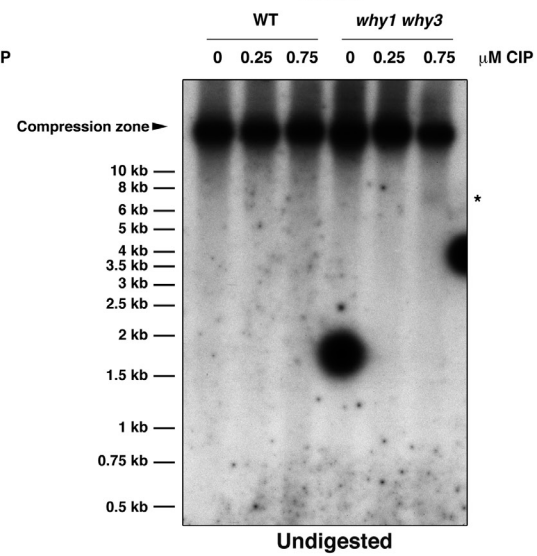
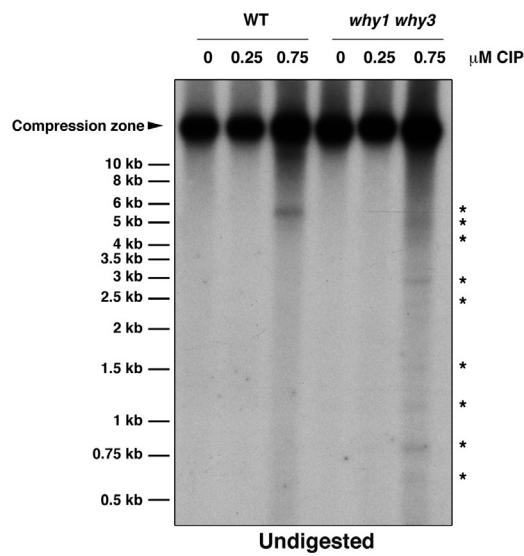
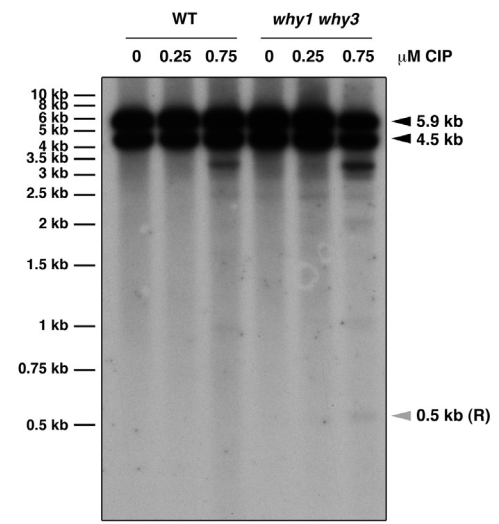
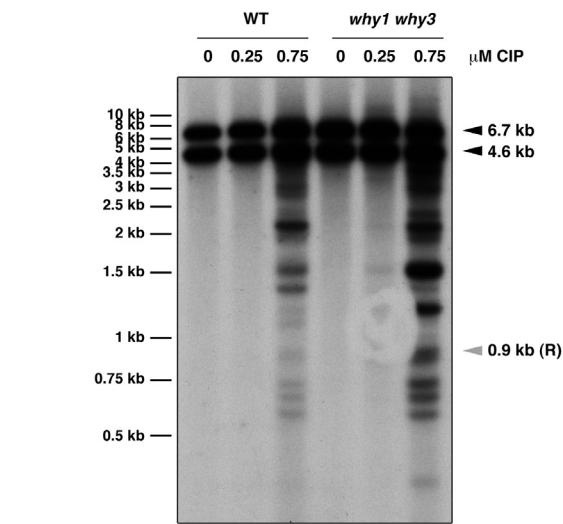
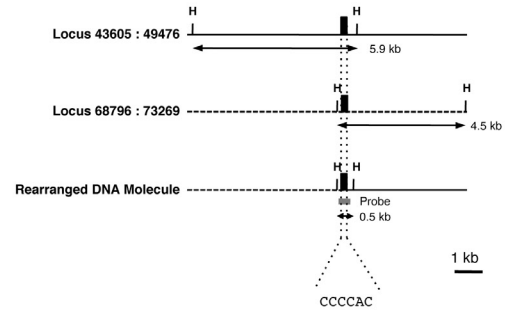
A**B**

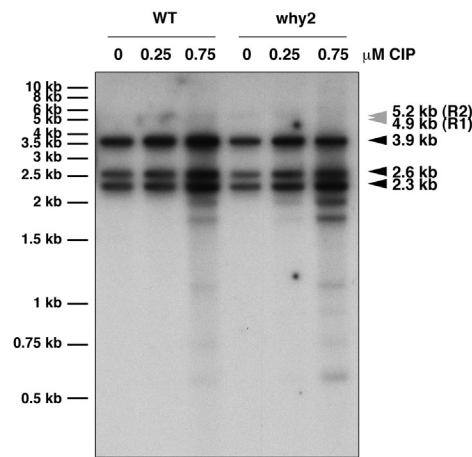
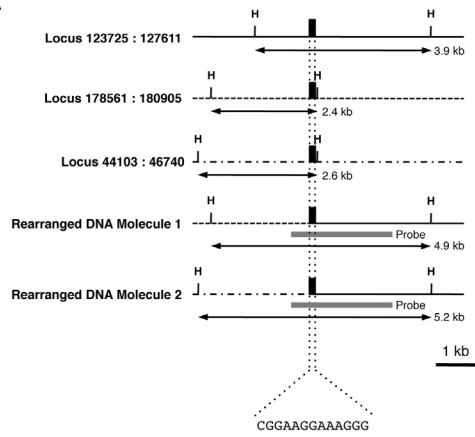
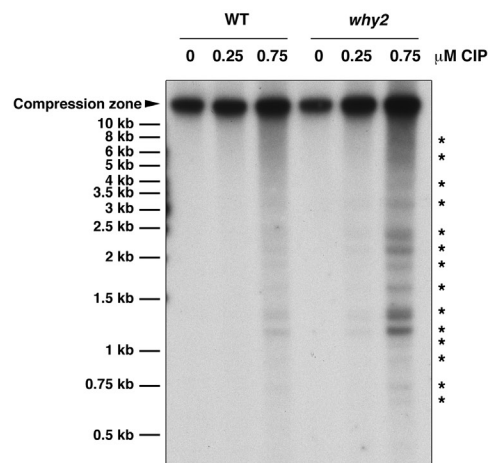
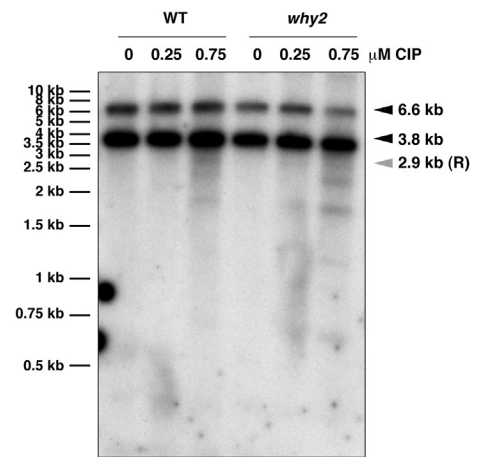
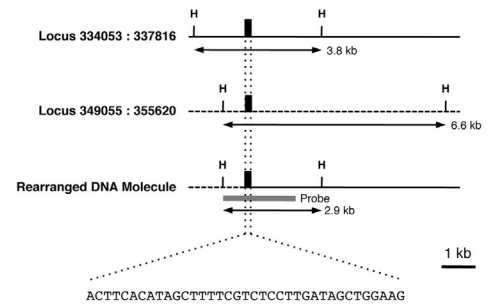
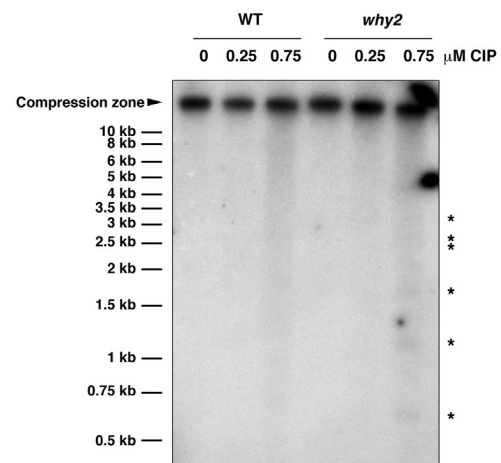
Figure 3-S4. DNA gel blot detection of DNA rearrangements in mitochondria following plant treatment with ciprofloxacin.

(A) Upper panel: Schematic representation of the microhomology-mediated DNA rearrangement occurring between two loci of the mitochondria genome of *Arabidopsis* (accession number NC_001284). The DNA rearrangement was isolated by PCR using the primers 125781REV and 180454FOR. The PCR product was used as a probe for the DNA gel blot presented in the middle and lower panels and is represented by a thick gray line. The HindIII fragments corresponding to the bands in the DNA gel blot in the middle panel are delimited by a “H”. The black boxes designate the microhomologous DNA sequences that mediated the rearrangement. The 2.3 and 2.6 kb bands both contain a 2 kb perfect repeat that could undergo DNA recombination. However, since a HindIII site interrupts the repeat, DNA recombination is not expected to generate new bands.

Middle panel: DNA gel blots of total DNA from WT and *why2-1* plants treated with ciprofloxacin. The total DNA was digested with HindIII and hybridized with the probe described in the upper panel. The predicted sizes of the bands corresponding to the WT and rearranged DNA molecules are indicated. R designates the rearranged DNA molecules.

Lower panel: Detection of subgenome-length molecules in DNA gel blots of total undigested DNA from WT and *why2-1* plants treated with ciprofloxacin using the same probe described in the upper panel. Asterisks indicate the position of rearranged DNA molecules.

(B) Same as in (A) but using as a probe a PCR product amplified with primers 187018REV and 349054FOR. The bands at 3.8 or 6.6 kb contain no repeats longer than 50 bp, suggesting that recombination cannot account for the apparition of new bands.

A**HindIII****Undigested****B****HindIII****Undigested**

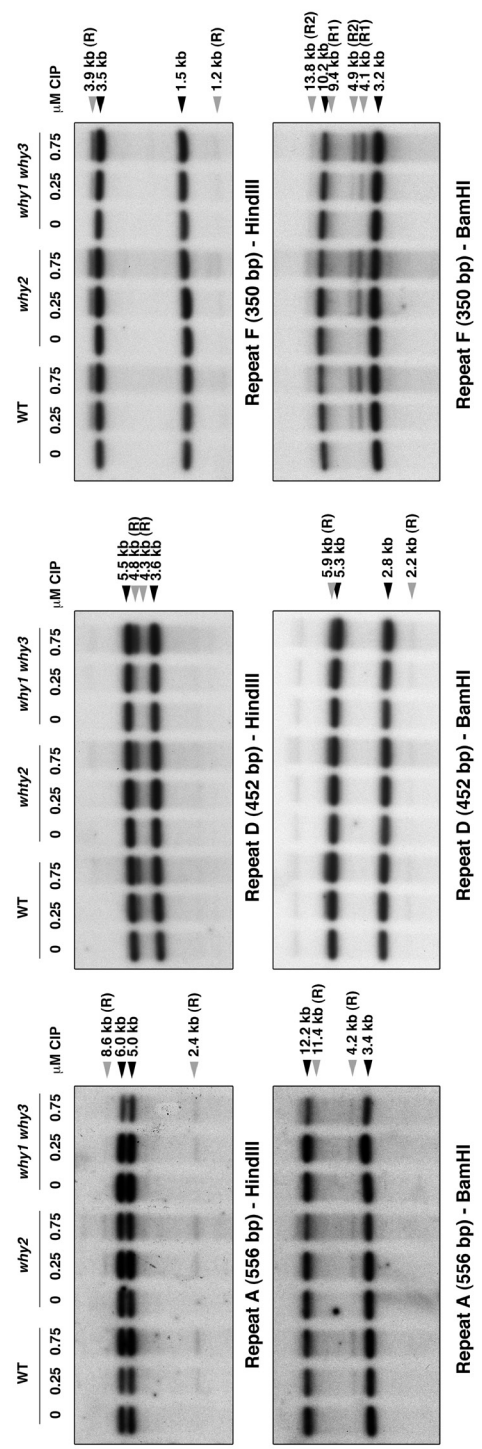


Figure 3-S5. Analysis of mitochondrial recombination in plants treated with ciprofloxacin. DNA gel blots of total DNA from WT, *why2-1* and *why1 why3* plants treated with ciprofloxacin. DNA gel blots were hybridized with probes encompassing mitochondrial repeats A, D and F. Total DNA was digested with HindIII or BamHI. Arrows indicate the expected parental forms (black arrow) and the recombination products (gray arrow). The predicted size of the expected bands are given. (R) designates a recombination product. In the case of the probe encompassing the repeat F, two recombination events can be detected upon BamHI digestion leading to the apparition of (R1) and (R2) recombination products. R1 recombination is mediated by the repeat F whereas R2 is mediated by a 206 bp repeat. The size of the repeats is given in parenthesis below each gel.

Figure 3-S7. Effects of bleomycin on plastid and mitochondria DNA rearrangements of plants lacking plastid- or mitochondria-targeted Whirlies.

(A) and (D) Phenotypic effects of various concentrations of bleomycin on (A) WT and *why1 why3* or (D) WT and *why2-1 Arabidopsis* plants. Plants were grown for three weeks on solid media containing the indicated concentrations of bleomycin.

(B) and (E) Electrophoretic analysis of representative PCR performed with 8 outward- or inward-facing (B) plastid genome-directed or (E) mitochondria genome-directed PCR primers on total leaf DNA of WT, *why2-1* and *why1 why3* plants treated with bleomycin. Low cycle amplification of the *ycf2* plastid gene or the *cox1* mitochondrial gene was used as a loading control. The oligonucleotides used for each PCR are indicated.

(C) and (F) Histograms showing the number of PCR products in the (C) plastid or (F) mitochondrial genome of WT, *why2-1* and *why1 why3* plants as a function of bleomycin concentration.

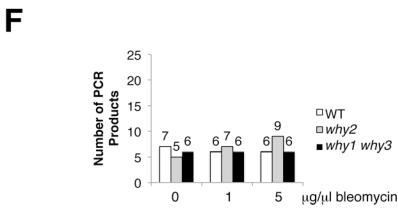
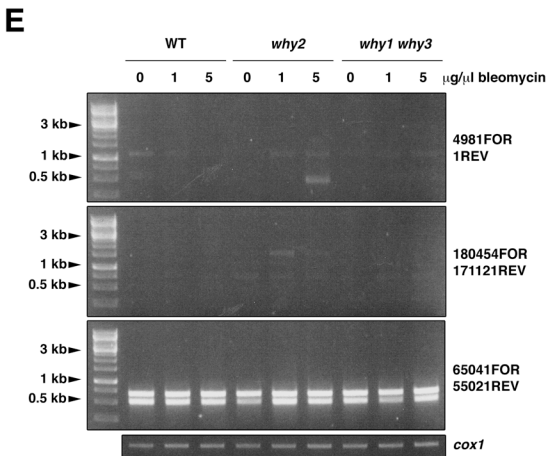
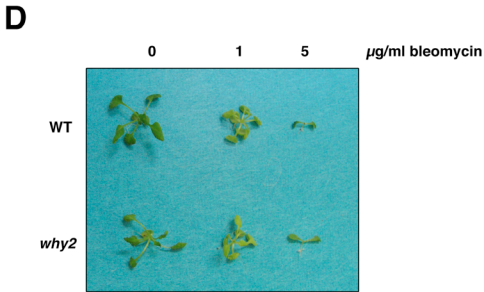
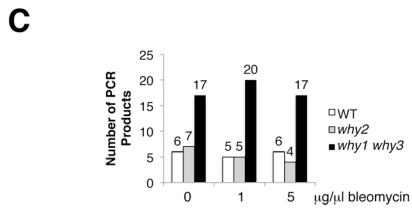
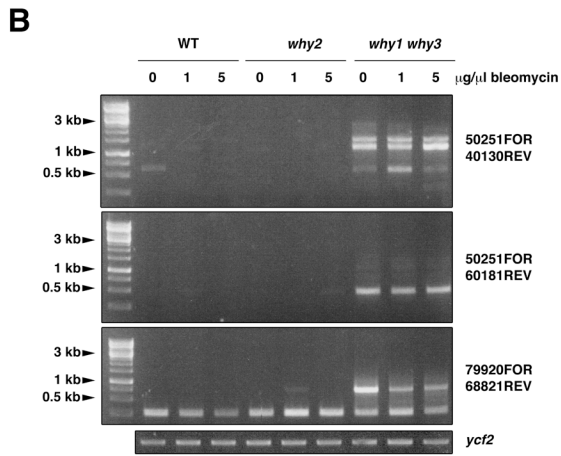
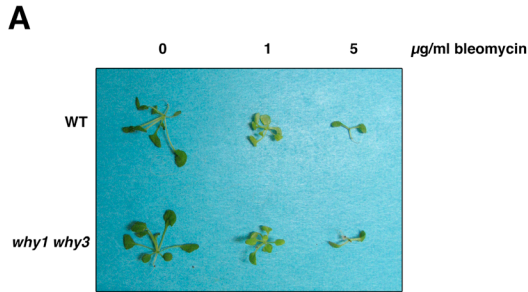
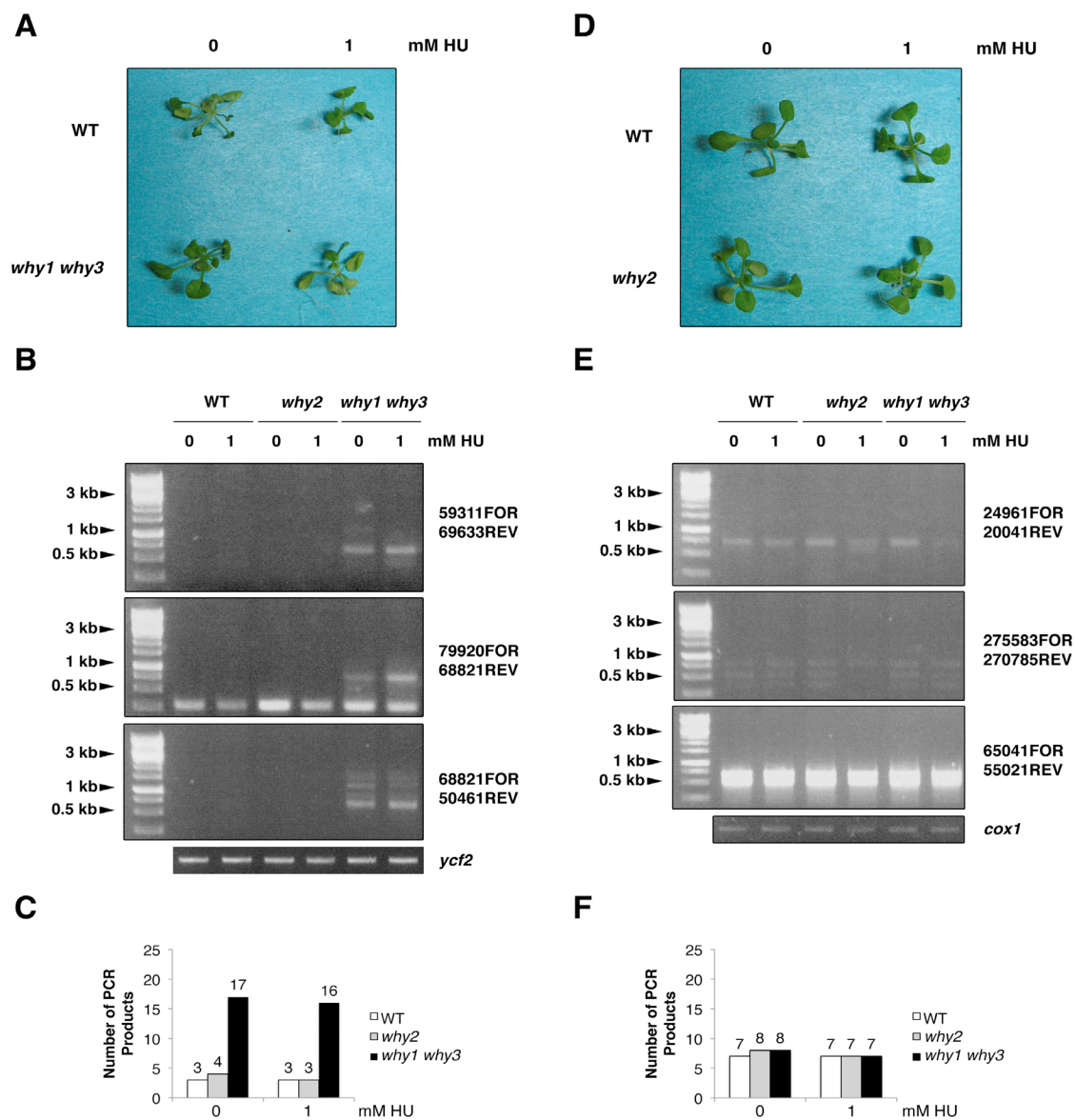


Figure 3-S8. Effects of hydroxyurea on plastid and mitochondria DNA rearrangements of plants lacking plastid- or mitochondria-targeted Whirlies.

(A) and (D) Phenotypic effects of various concentrations of hydroxyurea (HU) on (A) WT and *why1 why3* or (D) WT and *why2-1 Arabidopsis* plants. Plants were grown for three weeks on solid media containing the indicated concentrations of hydroxyurea.

(B) and (E) Electrophoretic analysis of representative PCR performed with 8 outward- or inward-facing (B) plastid genome-directed or (E) mitochondria genome-directed PCR primers on total leaf DNA of WT, *why2-1* and *why1 why3* plants treated with hydroxyurea. Low cycle amplification of the *ycf2* plastid gene or the *cox1* mitochondrial gene was used as a loading control. The oligonucleotides used for each PCR are indicated.

(C) and (F) Histograms showing the number of PCR products in the (C) plastid or (F) mitochondrial genome of WT, *why2-1* and *why1 why3* plants as a function of hydroxyurea concentration.



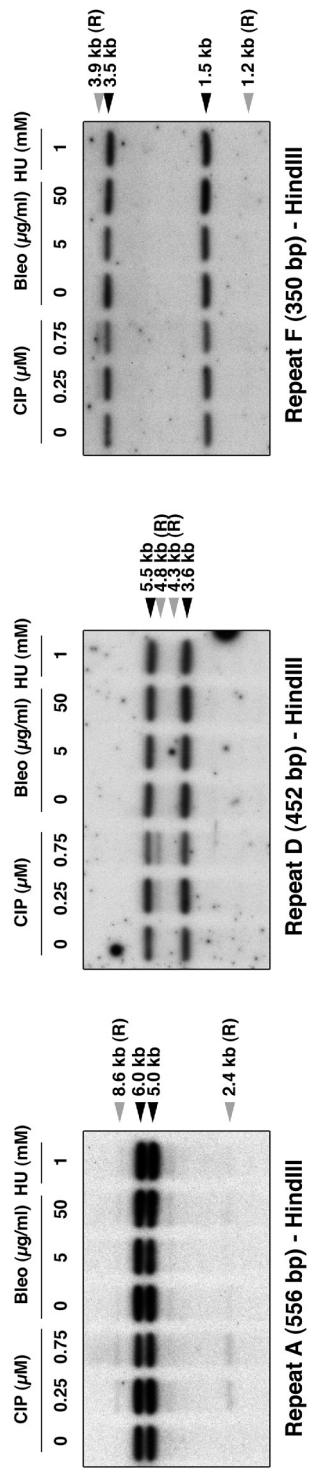


Figure 3-S9. Analysis of mitochondrial recombination in WT plants treated with bleomycin or hydroxyurea. DNA gel blots of total DNA from WT plants treated with ciprofloxacin, bleomycin or hydroxyurea. DNA gel blots were hybridized with probes encompassing mitochondrial repeats A, D and F. Total DNA was digested with HindIII. Arrows indicate the expected parental forms (black arrow) and the recombination products (gray arrow). The predicted sizes of the expected bands are given. (R) designates a recombination product. The size of the repeats is given in parenthesis below each gel.

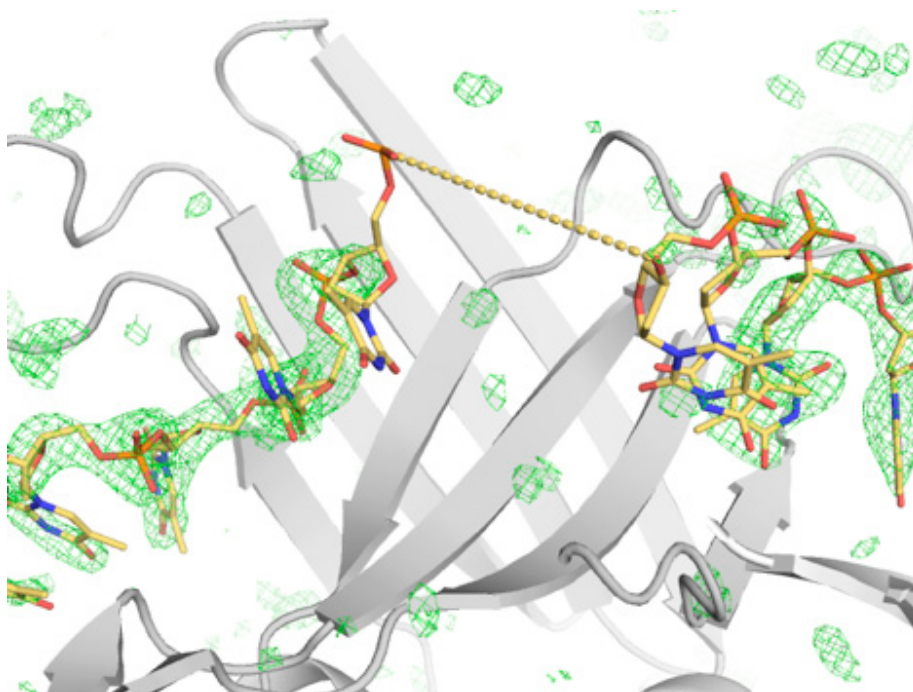


Figure 3-S10. Single-stranded DNA is stabilized on the edges of the β -sheets of Whirly proteins.

Crystal structure of St-WHY2 bound to ERE₃₂. St-WHY2 is in cartoon representation while DNA molecules are presented as stick models with carbon atoms colored in yellow. An $F_o - F_c$ simulated annealing omit map encompassing the entire DNA is contoured at 2.5σ and colored in green. A yellow dashed line indicates the possible path of ssDNA on top of the β -sheets.

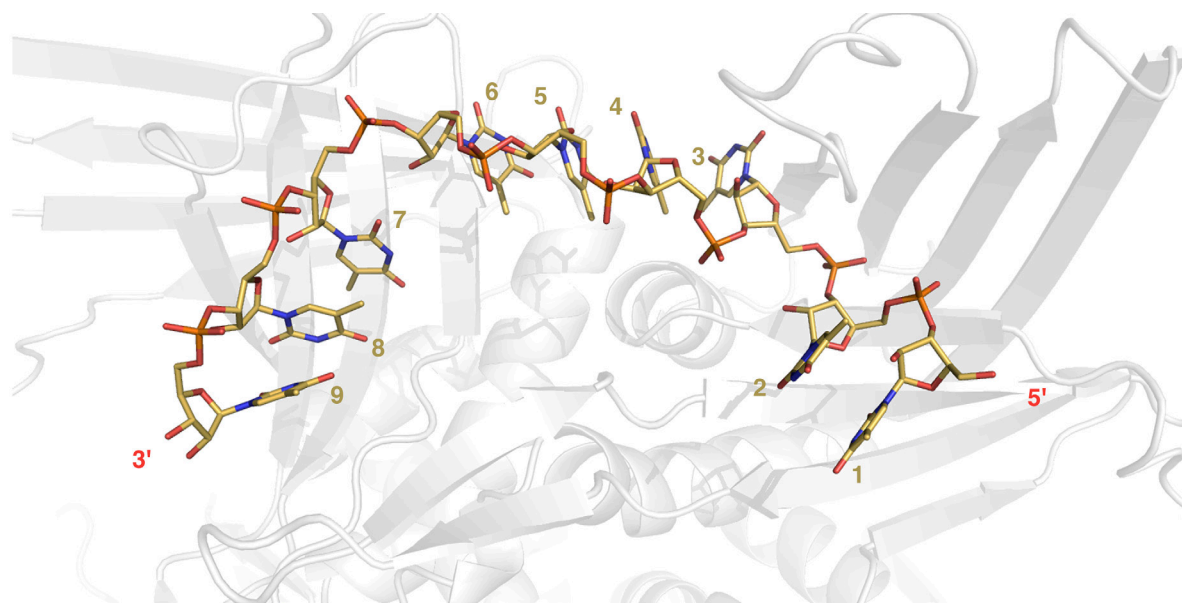


Figure 3-S11. Modeling of a St-WHY2-RNA complex.

The crystal structure of the St-WHY2-ERE₃₂ complex was used as a template to model RNA. The modeling was performed using CNS (Brünger et al., 1998) and Pymol (DeLano, 2002). The RNA is in stick representation with carbon atoms colored in yellow. The protein is in cartoon representation. Nucleotides residues are labeled.

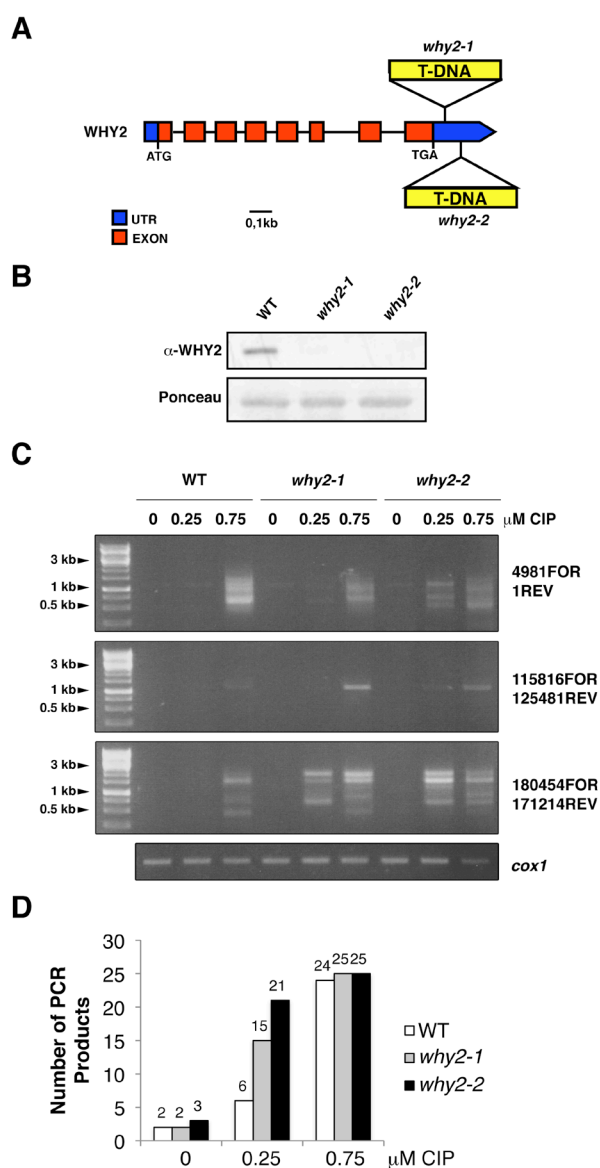
Figure 3-S12. Molecular characterization of *why2* lines.

(A) Physical map of the At-WHY2 (AT1G71260) gene. The positions of the T-DNA insertions in the *why2-1* and *why2-2* lines are indicated.

(B) WHY2 protein is undetectable in leaf protein extracts of *why2-1* and *why2-2* lines by protein gel blotting with an α -WHY2 antibody. A section of the blot stained with Ponceau red is presented below as a loading control.

(C) Electrophoretic analysis of representative PCR performed with 8 outward- or inward-facing mitochondria genome-directed PCR primers on total leaf DNA of WT, *why2-1* and *why2-2* plants treated with ciprofloxacin. Low cycle amplification of the *cox1* mitochondrial gene was used as a loading control. The oligonucleotides used for each PCR are indicated.

(D) Histogram showing the number of PCR products in WT, *why2-1* and *why2-2* plants as a function of ciprofloxacin concentration.



Event	Plant Treatment ^a	Plant Genotype	Junction Endpoints ^b	Junction Sequences ^c	Length of Junctions (Mismatches)	Junction Types ^d
Mitochondrial reaction 1 with primers 20041REV and 30001FOR ^{e,f}						
Mitochondrial reaction 2 with primers 30001REV and 50341FOR						
Mitochondrial reaction 3 with primers 55021REV and 65041FOR						
M3-1	CIP 0.75 μM	Col-0	54739 Recombinant 65547	aaaaaaaggg G aagaactac gtcgtgtac G aagaactac gtcgtgtac G ttcagtgtt	1 bp	NHEJ-like
M3-2	CIP 0.75 μM	Col-0	55001 Recombinant 65518	cgaaggagc cataccgaa tgagcgcgtg cataccgaa tgagcgcgtg tactgtgca	0 bp	NHEJ-like
M3-3	CIP 0.75 μM	why2	53421 Recombinant 65476	gtccgtcca TGGAGCGTATG attttccttg ctctaaagct TGGAGCGTATG attttccttg ctctaaagct TGGAGCGTATG gtaacggta	(0) 11 bp (0)	MHMR
M3-4	CIP 0.75 μM	why2	54638 Recombinant 65742	ccttattagc ctaatatctt atctcctct ctaatatctt atctcctct tcccttgact	0 bp	NHEJ-like
M3-5	CIP 0.75 μM	why2	54463 Recombinant 65476	tcgtgcctc gtcctaggct ctctaaagct gtcctaggct ctctaaagct tggagcgctat	0 bp	NHEJ-like
M3-6	CIP 0.75 μM	why2	54234 Recombinant 65096	taggaatctt AGGTTACCTCCAG cttcttgaag ccactagta AGGTTACCTCCAG cttcttgaag ccactagta AGGTTACTTCCAG agcgagtat	(0) 13 bp (1)	MHMR*
M3-7	CIP 0.75 μM	why2	54234 Recombinant 65096	taggaatctt AGGTTACCTCCAG cttcttgaag ccactagta AGGTTACTTCCAG cttcttgaag ccactagta AGGTTACTTCCAG agcgagtat	(1) 13 bp (0)	MHMR*
M3-8	CIP 0.75 μM	why2	54393 Recombinant 65074	cgcgacttgg GAAACCCCTT tcttcattcg ctgagctgta GAAACCCCTT tcttcattcg ctgagctgta GAAACCCCTT gatccactag	(0) 9 bp (0)	MHMR
M3-9	CIP 0.75 μM	why2	54483 Recombinant 65089	tgccttcgtc CTAGGTCGGCTACTTTC ttctcgaata ccttgatcca CTAGGTCGGCTACTTTC ttctcgaata ccttgatcca CTAGGTAAGGTTACTTTC cagacggagt	(0) 17 bp (4)	MHMR*
Mitochondrial reaction 4 with primers 130413REV and 151514FOR						
Mitochondrial reaction 5 with primers 180454REV and 200500FOR						
M5-1	CIP 0.75 μM	why2	179517 Recombinant 200771	atgacgggtt ACCCGG ttcaaggctt aaggtaaag ACCCGG ttcaaggctt aaggtaaag ACCCGG gaaagcgagg	(0) 6 bp (0)	MHMR
Mitochondrial reaction 6 with primers 235252REV and 245966FOR						
M6-1	CIP 0.25 μM	why2	234779 Recombinant 246725	gatgagatg AAGTAAAGGAATT actctggttg acgaaagtga AAGTAAAGGAATT actctggttg acgaaagtga AAGTAAAGGGAATT gtcggccct	(0) 13 bp (1)	MHMR*
Mitochondrial reaction 7 with primers 305346REV and 315513FOR ^{e,f}						

Table 3-SI (continued)

M7-1	why2	CIP 0.25 μM	305245 Recombinant 316714	cgaagataca AGCGCCCG gtcctcttct ttaagtccat AGCGCCCG gtcctcttct ttaagtccat AGCGCCCG attcgatcac	MHMR	(0) 8 bp (0)
M7-2	why2	CIP 0.25 μM	304953 Recombinant 316326	gataaaggta AGGTCACACT ctcaactagta taacaggcgt AGATCACACT gcaactagta taacaggcgt AGATCACACT ggaactggag	MHMR*	(1) 10 bp (0)
M7-3	why2	CIP 0.25 μM	305304 Recombinant 315835	aggttttgcg TCACAGAAGGAGTGG gaaattgggg aaagactacg TCACTGAAG-AGTGG gaaattgggg aaagactacg TCACTGAAG-AGTGG ttggaagac	MHMR*	(2) 14 bp (0)
M7-4	Col-0	CIP 0.75 μM	304556 Recombinant 316456	aaaagggaag AGGCTATTAAACATAAAAAAATAATAA agtccaaaatg atgcccctctc AGGCTAATGAAGCAAAAAAATAATAGAA agtccaaaatg atgcccctctc AGGCTAATGAAGCAAAAAAAGCGAA ggccttact	MHMR*	(6) 28 bp (2)
M7-5	Col-0	CIP 0.75 μM	305222 Recombinant 317005	cgataactcc tatcccgaag agctaggcag tatcccgaag agctaggcag atagatagac	NHEJ- like	0 bp
M7-6	Col-0	CIP 0.75 μM	305245 Recombinant 316714	cgaagataca AGCGCCCG gtcctcttct ttaagtccat AGCGCCCG gtcctcttct ttaagtccat AGCGCCCG attcgatcac	MHMR	(0) 8 bp (0)
M7-7	Col-0	CIP 0.75 μM	304900 Recombinant 316104	tggaatgctc TCCTTTGTTAAAAA ctacgagcgg caggggggat TCCTTTGTCGATAAAA ctacgagcgg cagggggggat TCCTTTGTCGATAGAA ggcattcgcg	MHMR*	(4) 15 bp (1)
M7-8	Col-0	CIP 0.75 μM	305105 Recombinant 316064	ccgagacctaa CGGCTA caagaataaa tagcaaacac CGGCTA caagaataaa tagcaaacac CGGCTA gataagtggg	MHMR	(0) 6 bp (0)
M7-9	why2	CIP 0.75 μM	305257 Recombinant 316731	cggtcctctt CTCT tgtgtcgaag gattcgatca CTCT tgtgtcgaag gattcgatca CTCT atcgaataga	NHEJ- like	4 bp
M7-10	why2	CIP 0.75 μM	305245 Recombinant 316714	cgaagataca AGCGCCCG gtcctcttct ttaagtccat AGCGCCCG gtcctcttct ttaagtccat AGCGCCCG attcgatcac	MHMR	(0) 8 bp (0)
M7-11	why2	CIP 0.75 μM	305252 Recombinant 316371	agatacaagc GCCTGGTCTCTT tctcttgtgt gacgaacaat GCTCGGCACCTT tctcttgtgt gacgaacaat GCTCGGCACCTT ggtcaagtga	MHMR*	(3) 12 bp (0)
M7-12	why2	CIP 0.75 μM	305132 Recombinant 315565	aaaagtacag GCAGAGCCAA accatcacta attcgatctc GCAGAGCCAA accatcacta attcgatctc GCAGAGCCAA ctggaacaga	MHMR*	(0) 10 bp (1)
Mitochondrial reaction 8 with primers 165904REV and 200500FOR						
Mitochondrial reaction 9 with primers 155034REV and 200500FOR						
Mitochondrial reaction 10 with primers 10081REV and 15001FOR						
M10-1	Col-0	CIP 0.75 μM	9505 Recombinant 16023	atcttagctt G agcctaactt aagtgaagg G agcctaactt aagtgaagg G cgctatccgg	NHEJ- like	1 bp
M10-2	why2	CIP 0.75 μM	9725 Recombinant 16003	ggttgtcact GCGAAGCTA-TAGTGAAGTGCG acgggggtctt tccttagagta GCGAAGCTACAAGTGAAGTGCG acgggggtctt tccttagagta GCGAAACTACAAGTGAAGGGCG gctatccgga	MHMR*	(2) 22 bp (2)
M10-3	why2	CIP 0.75 μM	9511 Recombinant 15480	gcttgacct A ccttcgcaca acgaagggtc A ccttcgcaca acgaagggtc A tccagaagcg	NHEJ- like	1 bp
Mitochondrial reaction 11 with primers 20461REV and 24961FOR						

Table 3-SI (continued)

M11-1	CIP 0.75 µM	Col-0	19795 Recombinant 25300	gaatccgagg ATTTCGTTGGATTTCGCGCT tcattctcccc atatcttaacc ATTTCGTTGGAAATTTCGCGCT tcattctcccc atatcttaacc ATTTCGTTGGAAATTTCGCGCT catagcacga	(4) 21 bp (0)	MHMR*
M11-2	CIP 0.75 µM	why2	19317 Recombinant 25309	caagacgata GATAGGCGG ttatgaaaa catttcttcg GAATTGCGG ttatgaaaa catttcttcg GAATTGCGG ttcatagcac	(5) 10 bp (1)	MHMR*
M11-3	CIP 0.75 µM	why2	19795 Recombinant 25300	gaatccgagg ATTTCGTTGGATTTCGCGCT tcattctcccc atatcttaacc ATTTCGTTGGAAATTTCGCGCT tcattctcccc atatcttaacc ATTTCGTTGGAAATTTCGCGCT catagcacga	(4) 21 bp (0)	MHMR*
Mitochondrial reaction 12 with primers 30001REV and 80161FOR						
M12-1	CIP 0.25 µM	Col-0	29609 Recombinant 80994	tttttccct CCCCCCCCCA aaaaaaaa attagacgtc CCCCCCCCCA aaaaaaaa attagacgtc CCCCCCCCCA tgctacggtt	(0) 10 bp (0)	MHMR
M12-2	CIP 0.25 µM	why2	29609 Recombinant 80994	tttttccct CCCCCCCCCA aaaaaaaa attagacgtc CCCCCCCCCA aaaaaaaa attagacgtc CCCCCCCCCA tgctacggtt	(0) 10 bp (0)	MHMR
M12-3	CIP 0.25 µM	why2	29610 Recombinant 80713	acttttttc CCT-CCCCCCCCCA aaaaaaaa cccgggcaaa CCT-CCCCCCCCCA aaaaaaaa cccgggcaaa CCAACCCCCCTCCAA gaggaattgt	(0) 14 bp (3)	MHMR*
M12-4	CIP 0.75 µM	Col-0	29609 Recombinant 80994	tttttccct CCCCCCCCCA aaaaaaaa attagacgtc CCCCCCCCCA aaaaaaaa attagacgtc CCCCCCCCCA tgctacggtt	(0) 10 bp (0)	MHMR
M12-5	CIP 0.75 µM	why2	29609 Recombinant 80994	tttttccct CCCCCCCCCA aaaaaaaa attagacgtc CCCCCCCCCA aaaaaaaa attagacgtc CCCCCCCCCA tgctacggtt	(0) 10 bp (0)	MHMR
M12-6	CIP 0.75 µM	why2	29610 Recombinant 80713	acttttttc CCT-CCCCCCCCCA aaaaaaaa cccgggcaaa CCT-CCCCCCCCCA aaaaaaaa cccgggcaaa CCAACCCCCCTCCAA gaggaattgt	(0) 14 bp (3)	MHMR*
Mitochondrial reaction 13 with primers 84901REV and 151514FOR¹						
Mitochondrial reaction 14 with primers 210239REV and 245966FOR¹						
Mitochondrial reaction 15 with primers 1REV and 4981FOR¹						
M15-1	CIP 0.25 µM	why2	366037 Recombinant 5462	ctaactgata ggccttggt ctttacctat ggccttggt ctttacctat cttgggtgag	0 bp	NHEJ- like
M15-2	CIP 0.25 µM	why2	366764 Recombinant 5145	gcgaatagaa GGCTT-TTGCCAC ctccctcaac gattctcttg GGCTTGTGCCAC ctccctcaac gattctcttg GGCTTGTGCCAC taaagaaaga	(1) 13 bp (0)	MHMR*
M15-3	CIP 0.75 µM	Col-0	366038 Recombinant 5171	taactgatag ggccttggt agaaagagt ggccttggt agaaagagt ctgccttcta	0 bp	NHEJ- like
M15-4	CIP 0.75 µM	Col-0	366042 Recombinant 5080	tgataggccc tggctggacc ccgtaggaaa tggctggacc ccgtaggaaa agcgtttgat	0 bp	NHEJ- like
M15-5	CIP 0.75 µM	Col-0	366617 Recombinant 5215	gccgacttta AGTCAAGGAGAAAGAGTG cctactgcca gcttccagc AGTCAAGGAGAAAGAGTG cctactgcca gcttccagc AGTCAACCAAGAAAGAGTG ttctccct	(0) 19 bp (3)	MHMR*

M15-6	CIP	why2	366041 Recombinant 5298	actgataggc C ctggctggac cacatctgct C ctggctggac cacatctgct C ttccaggac	1 bp	NHEJ- like
M15-7	CIP	why2	366856 Recombinant 5607	agaagaggtc TCAAGAGCC tcttgaattt aaactccat TCAAGGAGCC tcttgaattt aaactccat TCAAGGCGC aatcttcac	(1) 10 bp (1)	MHMR*
M15-8	CIP	why2	366856 Recombinant 5577	aggtctcaac GAGCC tcttgaattt ttttttataa GAGCC tcttgaattt ttttttataa GAGCC attgscaggga	(0) 5 bp (0)	MHMR
Mitochondrial reaction 16 with primers 125781REV and 115816FOR ¹ – Deletion Events						
M16-1	CIP	why2	125708 Recombinant 116865	taatgctttc TCTGCTTTCCCTCT tggtttcggt gggtgtcgag TCTGCTTTCCCTCT tggtttcggt gggtgtcgag TCTGCTTTCCCTCT ctatcaaacg	(0) 14 bp (0)	MHMR
M16-2	CIP	Col-0	125359 Recombinant 116280	gggtacgtcc CCTATC gcctttcatt gagcaggtaa CCTACC ttctaaacca	(1) 6 bp (0)	MHMR*
M16-3	CIP	why2	125529 Recombinant 116981	ttagctgagc ATCAATCCCAT tctttgcgga caggtatgaa ATCAATCCCAT tctttgcgga caggtatgaa ATCAATCCCAT gttgtgcgta	(0) 11 bp (0)	MHMR
M16-4	CIP	why2	125708 Recombinant 116865	taatgctttc TCTGCTTTCCCTCT tggtttcggt gggtgtcgag TCTGCTTTCCCTCT tggtttcggt gggtgtcgag TCTGCTTTCCCTCT ctatcaaacg	(0) 14 bp (0)	MHMR
Mitochondrial reaction 17 with primers 135180REV and 140053FOR ¹						
Mitochondrial reaction 18 with primers 171214REV and 180454FOR ¹						
M18-1	CIP	why2	170939 Recombinant 181121	attctataga TAQCTCTGGTAG gtaaacgact ggacaggtag TAGCTCTGGTAG gtaaacgact ggacaggtag TAGCTCTGGTAG agcgaggagc	(1) 12 bp (0)	MHMR*
M18-2	CIP	why2	171149 Recombinant 181066	aaccttcgtc TCAATCTTCATAA tcaattttatg tcacaaacta TCAATCTTCATAA tcaattttatg tcacaaacta TCAAT_TTCATAA gagaagacga	(0) 13 bp (1)	MHMR*
M18-3	CIP	Col-0	170678 Recombinant 181451	ggtttggatg gagctactgg attctccgtt gagctactgg attctccgtt tcgattattaa	0 bp	NHEJ- like
M18-4	CIP	Col-0	170609 Recombinant 180538	tttgtgatgc TCAACATGA-AACGATCGATATA gatcggcgta agtcacaaactg TGAACATGA-AACGATCGATATA gatcggcgta agtcacaaactg TGAACATGAAGAGATCATATA ttggtaaatag	(0) 22 bp (3)	MHMR*
M18-5	CIP	why2	170939 Recombinant 181121	attctataga TAQCTCTGGTAG gtaaacgact ggacaggtag TAGCTCTGGTAG gtaaacgact ggacaggtag TAGCTCTGGTAG agcgaggagc	(1) 12 bp (0)	MHMR*
M18-6	CIP	why2	171149 Recombinant 181066	aaccttcgtc TCAATCTTCATAA tcaattttatg tcacaaacta TCAATCTTCATAA tcaattttatg tcacaaacta TCAAT_TTCATAA gagaagacga	(0) 13 bp (1)	MHMR*
Mitochondrial reaction 19 with primers 245966REV and 239584FOR ¹ – Deletion Event						
M19-1	CIP	why2	244727 Recombinant 240924	gtaacccccc AAAAGGAAACTCACCCACCAAGA_ aaaagggtagg aaaaaaagaa AAAAGGAAACTCACCCACCAAGAAT aaaagggtagg aaaaaaagaa AAAAGGAAACTCACCCACCAAGAAT cagtccaata	(1) 25 bp (0)	MHMR*
Mitochondrial reaction 20 with primers 270785REV and 275583FOR ¹						

Table 3-SI (continued)

M20-1	CIP 0.25 μ M	<i>why2</i>	269741 Recombinant 276747	attgaagaaa <u>GGAGAAAGCCCTGTAAA</u> ctaccttact tccggtcatt GGTAGAAAGCCCTGTAAA ctaccttact tccggtcatt GGTAGAAAGCCCTGTAAA tggatttat	(1) 17 bp (0)	MHMR*
M20-2	CIP 0.25 μ M	<i>why2</i>	269069 Recombinant 275816	gtagtcccg CCGGATC aatagattgg tctcccaaac CCGGATC aatagattgg tctcccaaac CCGGATC cgcaagtgg	(0) 7 bp (0)	MHMR
M20-3	CIP 0.25 μ M	<i>why2</i>	270780 Recombinant 276407	tccaagccta ATAGGA acatcgctag tattccttcc ATAGGA acatcgctag tattccttcc ATAGGA ggacaccttg	(0) 6 bp (0)	MHMR
M20-4	CIP 0.75 μ M	Col-0	269800 Recombinant 276776	cagtaaaacta CCTCTCCTTCAGT gtttttccct gaatttatct CCTCTCCTTCAGT gtttttccct gaatttatct CCTCTCCTTCAGT cgagttccct	(1) 14 bp (0)	MHMR*
M20-5	CIP 0.75 μ M	Col-0	270679 Recombinant 276212	aggacggaac caccgaggaa tttgggtggt caccgaggaa tttgggtggt ttcacgaacc	0 bp	NHEJ- like
M20-6	CIP 0.75 μ M	Col-0	270497 Recombinant 276011	cattcatggc C ttttttggtt gagatgcgtg C ttttttggtt gagatgcgtg C gcacgggtga	1 bp	NHEJ- like
M20-7	CIP 0.75 μ M	Col-0	270716 Recombinant 275872	agtaggatgt atactgtgct tgctacctcc atactgtgct tgctacctcc caacagacgt	0 bp	NHEJ- like
M20-8	CIP 0.75 μ M	Col-0	270715 Recombinant 275738	tagtaggatg tatactgtgc ggaatatggc tatactgtgc ggaatatggc caccggatag	0 bp	NHEJ- like
M20-9	CIP 0.75 μ M	<i>why2</i>	269741 Recombinant 276747	attgaagaaa <u>GGAGAAAGCCCTGTAAA</u> ctaccttact tccggtcatt GGTAGAAAGCCCTGTAAA ctaccttact tccggtcatt GGTAGAAAGCCCTGTAAA tggatttat	(1) 17 bp (0)	MHMR*
M20-10	CIP 0.75 μ M	<i>why2</i>	270561 Recombinant 276533	ctgaggcgtt CCTCTCAGTCA tacaaaaaag atctctcag CCTCTCAGTCA tacaaaaaag atctctcag CCTCTCAGTCA aaagagtgcg	(0) 11 bp (0)	MHMR
M20-11	CIP 0.75 μ M	<i>why2</i>	270682 Recombinant 276415	ggacggaacc AC cgagggaatac ccataggagg AC cgagggaatac ccataggagg AC accttggttc	2 bp	NHEJ- like
M20-12	CIP 0.75 μ M	<i>why2</i>	270521 Recombinant 276184	ttctttgatt GATCTC cctttcgat ttctcttaa GATCTC cctttcgat ttctcttaa GATCTC aaccaaacac	(0) 6 bp (0)	MHMR
M20-13	CIP 0.75 μ M	<i>why2</i>	270135 Recombinant 275738	gatgtgaag CAC atcttagcaa ggaatatggc CAC atcttagcaa ggaatatggc CAC cggatagagc	3 bp	NHEJ- like
M20-14	CIP 0.75 μ M	<i>why2</i>	270445 Recombinant 275819	gcatttatca GATCAGC tcccagagta cccaaacccg GATCAGC tcccagagta cccaaacccg GATCAGC aagtcgggtc	(0) 7 bp (1)	MHMR*
M20-15	CIP 0.75 μ M	<i>why2</i>	270679 Recombinant 275979	aggacggaac caccgaggaa agaaagggtcg caccgaggaa agaaagggtcg tatcggtcag	0 bp	NHEJ- like

Plastid reaction 1 with primers 31138REV and 19181FOR¹ – Deletion Events

Table 3-SI (continued)

P1-1	CIP 0.25 μ M	Col-0	30922 Recombinant 20410	aaatagtagt atccctgtct atgatagtta atccctgtct atgatagtta ccaaaagatt	0 bp	NHEJ- like
P1-2	CIP 0.25 μ M	Col-0	31024 Recombinant 20347	gttaagttat CAAAAATTTGCTTTTTT ttttttagaa aatatatata CAAAAATTTTCTTTTTTT ttttttagaa aatatatata CAAAAATTTTCTTTTTTT acacttctta	(1) 18 bp (0)	MHMR*
P1-3	CIP 0.25 μ M	why1 why3	31024 Recombinant 20347	gttaagttat CAAAAATTTGCTTTTTTT ttttttagaa aatatatata CAAAAATTTTCTTTTTTT ttttttagaa aatatatata CAAAAATTTTCTTTTTTT acacttctta	(1) 18 bp (0)	MHMR*
P1-4	CIP 0.75 μ M	why1 why3	31024 Recombinant 20347	gttaagttat CAAAAATTTGCTTTTTTT ttttttagaa aatatatata CAAAAATTTTCTTTTTTT ttttttagaa aatatatata CAAAAATTTTCTTTTTTT acacttctta	(1) 18 bp (0)	MHMR*
P1-5	NOV 0 μ M	why1 why3	31024 Recombinant 20347	gttaagttat CAAAAATTTGCTTTTTTT ttttttagaa aatatatata CAAAAATTTTCTTTTTTT ttttttagaa aatatatata CAAAAATTTTCTTTTTTT acacttctta	(1) 18 bp (0)	MHMR*
P1-6	NOV 100 μ M	why1 why3	31024 Recombinant 20347	gttaagttat CAAAAATTTGCTTTTTTT ttttttagaa aatatatata CAAAAATTTTCTTTTTTT ttttttagaa aatatatata CAAAAATTTTCTTTTTTT acacttctta	(1) 18 bp (0)	MHMR*
P1-7	NOV 100 μ M	why1 why3	30618 Recombinant 19900	tataatgaaa TGAATGG tatgactagc caattcctaa TGAATGG tatgactagc caattcctaa TGAATGG atgtacgact	(0) 8 bp (0)	MHMR
Plastid reaction 2 with primers 40130REV and 50251FOR ¹						
P2-1	CIP 0 μ M	Col-0	39586 Recombinant 50325	ttaatgcat AATGCC ttccaatcc tattagttaa AATGCC ttccaatcc tattagttaa AATGCC cagaaatat	(0) 6 bp (0)	MHMR
P2-2	CIP 0 μ M	why1 why3	38759 Recombinant 50687	ttgcgatgc TAAATGATGATGT gccatctcgg cttcgttttt TAAATGATGATGT gccatctcgg cttcgttttt TAAATGATGATGT atcaaaaaat	(0) 13 bp (1)	MHMR*
P2-3	CIP 0 μ M	why1 why3	38989 Recombinant 50299	ccctggatg CAGGAATAGC gacatggact aatgccaaaa CAGGAATAGC gacatggact aatgccaaaa CAGGAATAGC acttgatatt	(0) 10 bp (0)	MHMR
P2-4	CIP 0.25 μ M	Col-0	39586 Recombinant 50325	ttaatgcat AATGCC ttccaatcc tattagttaa AATGCC ttccaatcc tattagttaa AATGCC cagaaatat	(0) 6 bp (0)	MHMR
P2-5	CIP 0.25 μ M	why1 why3	39109 Recombinant 50758	cttggtttcc ATTTTGGTTG taggtgtaac gatagattga ATTTTGGTTG taggtgtaac gatagattga ATTTTGGTTG aatttaatta	(0) 10 bp (0)	MHMR
P2-6	CIP 0.25 μ M	why1 why3	38989 Recombinant 50299	ccctggatg CAGGAATAGC gacatggact aatgccaaaa CAGGAATAGC gacatggact aatgccaaaa CAGGAATAGC acttgatatt	(0) 10 bp (0)	MHMR
P2-7	CIP 0.25 μ M	why1 why3	39517 Recombinant 50285	tcatgtgcg TAGCAATACCAACCA aatacgaaga ctccggaat TAGCAATACCAACCA aatacgaaga ctccggaat TAGAATGCCAAACA ggaatagcac	(0) 16 bp (3)	MHMR*
P2-8	CIP 0.25 μ M	why1 why3	39582 Recombinant 50289	tggaaatctt AATGCCATAA tgcctttcaa ggaaattaga AATGCCATAA tgcctttcaa ggaaattaga AATGCCAAA caggaatagc	(0) 10 bp (1)	MHMR*
P2-9	CIP 0.75 μ M	Col-0	37846 Recombinant 50396	ccagaaact GC caataaaat aaaaatccc GC caataaaat aaaaatccc GC ttagtcaatt	2 bp	NHEJ- like

Table 3-SI (continued)

P3-1	CIP 0.25 µM	why1 why3	83151 Recombinant 95532	ggaattctg GTAATAGCAATG ttgagtttcc gctgtttga GTAATAGCAATG ttgagtttcc gctgtttga GTAATAGCAATG agattcccgga	(0) 12 bp (0)	MHMR
P3-2	CIP 0.25 µM	why1 why3	142740 Recombinant 36	aaatcagtc ATTCACAA-CCACTG atgaaggatt gcgatggga ATTCACAATCCACTG atgaaggatt gcgatggga ATTCACAATCCACTG ccttaatacca	(1) 15 bp (0)	MHMR*
P3-3	CIP 0.25 µM	why1 why3	84368 Recombinant 95699	acgcgcagca AGAATCAAAGTC tccctatatt ctggagtggg AGAATCAAAGTC tccctatatt ctggagtggg AGA TC CAGATC gaaaagagga	(0) 12 bp (2)	MHMR*
P3-4	CIP 0.25 µM	why1 why3	84661 Recombinant 95481	tttaattgg CCGATTTT cccctctttt aattacatat CCGATTTT cccctctttt aattacatat CCGATTTT gaocctatgga	(0) 7 bp (0)	MHMR
P3-5	CIP 0.75 µM	Col-0	84464 Recombinant 96061	gtcgacgggat gctcctattta gtcgacgggat ccaacctcat cctcccttca ccaacctcat	0 bp	NHEJ- like
P3-6	CIP 0.75 µM	why1 why3	83151 Recombinant 95532	ggaattctg GTAATAGCAATG ttgagtttcc gctgtttga GTAATAGCAATG ttgagtttcc gctgtttga GTAATAGCAATG agattcccgga	(0) 12 bp (0)	MHMR
P3-7	CIP 0.75 µM	why1 why3	83532 Recombinant 95421	ttgattttct GTTTTCAT tgcctgcacc cagcatataa GTTTTCAT tgcctgcacc cagcatataa GTTATCAT gctcgcatat	(0) 8 bp (1)	MHMR*
P3-8	CIP 0.75 µM	why1 why3	142740 Recombinant 36	aaatcagtc ATTCACAA-CCACTG atgaaggatt gcgatggga ATTCACAATCCACTG atgaaggatt gcgatggga ATTCACAATCCACTG ccttaatacca	(1) 15 bp (0)	MHMR*
P3-9	CIP 0.75 µM	why1 why3	143168 Recombinant 15398	tccataggtc AAATCGG atatgttaatt aaaagagggg AAATCGG atatgttaatt aaaagagggg AAATCGG ccacattaaa	(0) 7 bp (0)	MHMR
Plastid reaction 4 with primers 105961REV and 116007FOR ¹						
P4-1	CIP 0.25 µM	why1 why3	105738 Recombinant 116787	agctaaagcc CCTAAATGACCGCT cagtgataaaa agaaaatgat CCTAAATGACCGCT cagtgataaaa agaaaatgat OCTATTTGACCGCT gtacattgct	(0) 14 bp (2)	MHMR*
P4-2	CIP 0.25 µM	why1 why3	105452 Recombinant 116276	ggaggtccga ACCGACTGATGTTGAA aaatcagcgg aattacgttg ACCGGGAGATGTTGAA aaatcagcgg aattacgttg ACCGGGAGATGTTGAA gctgcataaaa	(3) 16 bp (0)	MHMR*
P4-3	CIP 0.25 µM	why1 why3	105322 Recombinant 116009	tggtcttggtt AAGGGAACCCACG gagccgtagc ccggcaatgc AAGGGAAGCCATCG gagccgtagc ccggcaatgc AAGGGAAGCCATCG ataaaatagt	(2) 14 bp (0)	MHMR*
P4-4	CIP 0.75 µM	Col-0	105739 Recombinant 116394	cctaaatgacc GCTC agtgataaag caaccccatag GCTC agtgataaag caaccccatag GCTC ccatttttaa	4 bp	NHEJ- like
P4-5	CIP 0.75 µM	why1 why3	105452 Recombinant 116276	ggaggtccga ACCGACTGATGTTGAA aaatcagcgg aattacgttg ACCGGGAGATGTTGAA aaatcagcgg aattacgttg ACCGGGAGATGTTGAA gctgcataaaa	(3) 16 bp (0)	MHMR*
P4-6	CIP 0.75 µM	why1 why3	105821 Recombinant 116393	gtgcgtaata GCTC actgatacag caaccccatag GCTC actgatacag caaccccatag GCTC ccatttttaa	4 bp	NHEJ- like
Plastid reaction 5 with primers 69633REV and 59291FOR ¹ – Deletion Events						

Table 3-SI (continued)

P5-1	CIP 0.25 μ M	<i>why1</i> <i>why3</i>	68984 Recombinant 59308	tctttccggc AATTGCAATGGCTT ctttatttct gtataaaaa AATTCCAATGGCTT ttgtactct	(1) 14 bp (0)	MHMR*
P5-2	CIP 0.75 μ M	<i>why1</i> <i>why3</i>	68984 Recombinant 59308	tctttccggc AATTGCAATGGCTT ctttatttct gtataaaaa AATTCCAATGGCTT ttgtactct	(1) 14 bp (0)	MHMR*
P5-3	NOV 0 μ M	<i>why1</i> <i>why3</i>	68522 Recombinant 59380	atttcttcca ATCCCTTATTTTTT atgtcattcg accgaatcat ATCACCCTTTTTTTTT tattttaaaa	(2) 16 bp (0)	MHMR*
P5-4	NOV 100 μ M	<i>why1</i> <i>why3</i>	68984 Recombinant 59308	tctttccggc AATTGCAATGGCTT ctttatttct gtataaaaa AATTCCAATGGCTT ttgtactct	(1) 14 bp (0)	MHMR*
Plastid reaction 6 with primers 111330REV and 121141FOR^{e1}						
P6-1	CIP 0 μ M	<i>why1</i> <i>why3</i>	109932 Recombinant 121745	atgcaacaac AGATGTTATTTGTAAACAAGTAGTTTTGT tggttgggta tagatggacg AAGATTTCTTTTG-AA-AAGTAGTTTTGT tggttgggta tagatggacg AAGATTTCTTTTG-AA-AAGTAGTTTTGT cccatcggt	(4) 27 bp (0)	MHMR*
P6-2	CIP 0 μ M	<i>why1</i> <i>why3</i>	111043 Recombinant 12213	cggataatag AAAGTTCCTGTATT aagtcacaaa actcttacat AAAGTTCCTGTATT aagtcacaaa actcttacat AAAGTTCCTGTATT gataattcaa	(0) 14 bp (1)	MHMR*
P6-3a	CIP 0 μ M	<i>why1</i> <i>why3</i>	19786 Recombinant 110850	catatctcaa TAGATTGGCGT-AATTTTT ctaccgcacg caaaatttgg TAGATTGGCGT-AATTTTT ctaccgcacg caaaatttgg TAGATTGGTATGAATTTT gagaatgca	(0) 18 bp (3)	MHMR*
P6-3b	CIP 0 μ M	<i>why1</i> <i>why3</i>	19867 Recombinant 122003	taattatagg AGAAGTTGATCGTTAAATGAT aatttatgca gccagccctt AGAAGTTGATCGTTAAATGAT aatttatgca gccagccctt AGAAGTT-ATCGTTAAAAA-GAT catcaattcc	(0) 22 bp (3)	MHMR*
P6-4	CIP 0 μ M	<i>why1</i> <i>why3</i>	111244 Recombinant 121304	tagaaggagc CAGTTTTTTTTTCCACT gtaattttaga aatttcagag CAGTTTTTTTTTCCACT gtaattttaga aatttcagag CAGTTTTTTTTTCCACT catataaac	(2) 16 bp (0)	MHMR*
P6-5	CIP 0.25 μ M	Col-0	109932 Recombinant 121745	atgcaacaac AAGATGTTATTTGTAAACAAGTAGTTTTGT tggttgggta tagatggacg AAGATTTCTTTTG-AA-AAGTAGTTTTGT tggttgggta tagatggacg AAGATTTCTTTTG-AA-AAGTAGTTTTGT cccatcggt	(4) 27 bp (0)	MHMR*
P6-6	CIP 0.25 μ M	Col-0	109826 Recombinant 121273	cctgaataat CTCATTTTTT aattattcaa ggaaatgaga CTCATTTTTT aattattcaa ggaaatgaga CTCATTTTTT gttttactgc	(0) 10 bp (0)	MHMR
P6-7	CIP 0.25 μ M	Col-0	111043 Recombinant 12213	cggataatag AAAGTTCCTGTATT aagtcacaaa actcttacat AAAGTTCCTGTATT aagtcacaaa actcttacat AAAGTTCCTGTATT gataattcaa	(0) 14 bp (1)	MHMR*
P6-8	CIP 0.25 μ M	Col-0	111244 Recombinant 121304	tagaaggagc CAGTTTTTTTTTCCACT gtaattttaga aatttcagag CAGTTTTTTTTTCCACT gtaattttaga aatttcagag CAGTTTTTTTTTCCACT catataaac	(2) 16 bp (0)	MHMR*
P6-9	CIP 0.25 μ M	<i>why1</i> <i>why3</i>	109932 Recombinant 121745	atgcaacaac AAGATGTTATTTGTAAACAAGTAGTTTTGT tggttgggta tagatggacg AAGATTTCTTTTG-AA-AAGTAGTTTTGT tggttgggta tagatggacg AAGATTTCTTTTG-AA-AAGTAGTTTTGT cccatcggt	(4) 27 bp (0)	MHMR*
P6-10	CIP 0.25 μ M	<i>why1</i> <i>why3</i>	111244 Recombinant 121304	tagaaggagc CAGTTTTTTTTTCCACT gtaattttaga aatttcagag CAGTTTTTTTTTCCACT gtaattttaga aatttcagag CAGTTTTTTTTTCCACT catataaac	(2) 16 bp (0)	MHMR*
P6-11	CIP 0.75 μ M	Col-0	111043 Recombinant 12213	cggataatag AAAGTTCCTGTATT aagtcacaaa actcttacat AAAGTTCCTGTATT aagtcacaaa actcttacat AAAGTTCCTGTATT gataattcaa	(0) 14 bp (1)	MHMR*

Table 3-SI (continued)

P6-12	CIP 0.75 μ M	Col-0	111244 Recombinant 121304	tagaaggagc CAGTTTTTTTCCACT gtaatttaga aatttcagag CAGTTTTTTTTCCT gtaatttaga aatttcagag CAGTTTTTTTTCCT catataaac	(2) 16 bp (0)	MHMR*
P6-13	CIP 0.75 μ M	why1 why3	109932 Recombinant 121745	atgcaacaac AAGATTTTATTTGTAACAGTAGTTTGT tgggttggtta tagatggagc AAGATTTTCTTTG-AA-AAGTAGTTTGT tgggttggtta tagatggagc AAGATTTTCTTTG-AA-AAGTAGTTTGT cccatcggt	(4) 27 bp (0)	MHMR*
P6-14	CIP 0.75 μ M	why1 why3	111244 Recombinant 121304	tagaaggagc CAGTTTTTTTCCACT gtaatttaga aatttcagag CAGTTTTTTTTCCT gtaatttaga aatttcagag CAGTTTTTTTTCCT catataaac	(2) 16 bp (0)	MHMR*
P6-15	NOV 0 μ M	why1 why3	111189 Recombinant 121305	atttttatta AGTTTTTTTT ctctctcttt atttcagagc AGTTTTTTTT ctctctcttt atttcagagc AGTTTTTTTT tcaactcatat	(0) 10 bp (0)	MHMR
Plastid reaction 7 with primers 50513REV and 59311FOR¹						
P7-1	CIP 0.25 μ M	why1 why3	50489 Recombinant 59938	aggttaatcg AATCATTTTTTTTC gtttggttgt cgctccagca AATCATTTTTTTTC gtttggttgt cgctccagca AATCATTTTTTTTC cacaaggat	(0) 14 bp (0)	MHMR
P7-2	CIP 0.75 μ M	Col-0	50489 Recombinant 59938	aggttaatcg AATCATTTTTTTTC gtttggttgt cgctccagca AATCATTTTTTTTC gtttggttgt cgctccagca AATCATTTTTTTTC cacaaggat	(0) 14 bp (0)	MHMR
P7-3	CIP 0.75 μ M	why1 why3	50489 Recombinant 59938	aggttaatcg AATCATTTTTTTTC gtttggttgt cgctccagca AATCATTTTTTTTC gtttggttgt cgctccagca AATCATTTTTTTTC cacaaggat	(0) 14 bp (0)	MHMR
P7-4	CIP 0.75 μ M	why1 why3	50462 Recombinant 59705	ctcttgatc AAAACAAAAA ticaaggttaa agttactctg AAAACAAAAA ticaaggttaa agttactctg AAAACAAAAA aaaggaaagt	(0) 10 bp (0)	MHMR
P7-5	NOV 0 μ M	why1 why3	50489 Recombinant 59938	aggttaatcg AATCATTTTTTTTC gtttggttgt cgctccagca AATCATTTTTTTTC gtttggttgt cgctccagca AATCATTTTTTTTC cacaaggat	(0) 14 bp (0)	MHMR
Plastid reaction 8 with primers 68821REV and 79920FOR¹						
P8-1	CIP 0 μ M	why1 why3	68809 Recombinant 81088	cctcgaaagc TTGATGCAAA taaacgaagt atctctctt TTAATGCAAA taaacgaagt atctctctt TTAATGCAAA gacgaagta	(1) 10 bp (0)	MHMR*
P8-2	CIP 0 μ M	why1 why3	68714 Recombinant 80522	ggcgaaatc TCTTTTCTTTTA cccctatecc atcaacatgt TCTTTTCTTTTA cccctatecc atcaacatgt TCTTTTCTTTTA ttatatata	(0) 13 bp (0)	MHMR
P8-3	CIP 0.25 μ M	why1 why3	68714 Recombinant 80522	ggcgaaatc TCTTTTCTTTTA cccctatecc atcaacatgt TCTTTTCTTTTA cccctatecc atcaacatgt TCTTTTCTTTTA ttatatata	(0) 13 bp (0)	MHMR
P8-4	CIP 0.75 μ M	Col-0	68733 Recombinant 81329	cogatgagc G aaactaaagc ataactcagg G aaactaaagc ataactcagg G gatactttcc	1 bp	NHEJ- like
P8-5	CIP 0.75 μ M	Col-0	68714 Recombinant 80522	ggcgaaatc TCTTTTCTTTTA cccctatecc atcaacatgt TCTTTTCTTTTA cccctatecc atcaacatgt TCTTTTCTTTTA ttatatata	(0) 13 bp (0)	MHMR
P8-6	CIP 0.75 μ M	why1 why3	68733 Recombinant 82098	cogatgagc G aaactaaagc tatattcatg G aaactaaagc tatattcatg G gtccgctgt	1 bp	NHEJ- like

Table 3-SI (continued)

P8-7	CIP 0.75 μM	<i>why1</i> <i>why3</i>	68714 Recombinant 80522	ggcgaatac TCCTTTTCCTTTTA cccctatccc atcaacatgt TCCTTTTCCTTTTA cccctatccc atcaacatgt TCCTTTTCCTTTTA ttatatata	(0) 13 bp (0)	MHMR
P8-8	NOV 0 μM	<i>why1</i> <i>why3</i>	68522 Recombinant 80652	atcccttat TTTTTC atgtcatcgc tcaataacctg TTTTTC atgtcatcgc tcaataacctg TTTTTC taatccgata	(0) 6 bp (0)	MHMR
P8-9	NOV 0 μM	<i>why1</i> <i>why3</i>	68714 Recombinant 80522	ggcgaatac TCCTTTTCCTTTTA cccctatccc atcaacatgt TCCTTTTCCTTTTA cccctatccc atcaacatgt TCCTTTTCCTTTTA ttatatata	(0) 13 bp (0)	MHMR
P8-10	NOV 100 μM	Col-0	68714 Recombinant 80522	ggcgaatac TCCTTTTCCTTTTA cccctatccc atcaacatgt TCCTTTTCCTTTTA cccctatccc atcaacatgt TCCTTTTCCTTTTA ttatatata	(0) 13 bp (0)	MHMR
P8-11	NOV 100 μM	<i>why1</i> <i>why3</i>	68819 Recombinant 81100	atgcaataaa ACGAAGT ttgtttctac aatgcaaaagg ACGAAGT ttgtttctac aatgcaaaagg ACGAAGT aaaaaaat	(0) 7 bp (0)	MHMR
P8-12	NOV 100 μM	<i>why1</i> <i>why3</i>	68714 Recombinant 80522	ggcgaatac TCCTTTTCCTTTTA cccctatccc atcaacatgt TCCTTTTCCTTTTA cccctatccc atcaacatgt TCCTTTTCCTTTTA ttatatata	(0) 13 bp (0)	MHMR
Plastid reaction 9 with primers 10230REV and 30511FOR						
P9-1	CIP 0 μM	<i>why1</i> <i>why3</i>	9940 Recombinant 30859	agaataatcc GTTTCGTTTTTTT ataattcttc agaggggttaa GTTTCGTTTTTTT ataattcttc agaggggttaa GTTTCGTTTTTTT ttacttctc	(0) 12 bp (0)	MHMR
P9-2	CIP 0 μM	<i>why1</i> <i>why3</i>	9948 Recombinant 30859	agaataatcc GTTTCGTTTTTTT--ATACTTTC tctgaagta agaggggttaa GTTTCGTTTTTTTACTTTC tctgaagta agaggggttaa GTTTCGTTTTTTTACTTTC cttgatagat	(3) 22 bp (0)	MHMR*
P9-3	CIP 0 μM	<i>why1</i> <i>why3</i>	9895 Recombinant 30750	taaaaaata ATTTAAATTAAGAGTAAA attgactaaa aaatggaat ATTTAAATTAAGAGTAAA attgactaaa aaatggaat ATTTAAATTAAGAGTAAA taaaaaaaa	(0) 19 bp (2)	MHMR*
P9-4	CIP 0.25 μM	Col-0	9948 Recombinant 30859	agaataatcc GTTTCGTTTTTTT--ATACTTTC tctgaagta agaggggttaa GTTTCGTTTTTTTACTTTC tctgaagta agaggggttaa GTTTCGTTTTTTTACTTTC cttgatagat	(3) 22 bp (0)	MHMR*
P9-5	CIP 0.25 μM	<i>why1</i> <i>why3</i>	9476 Recombinant 31012	aaattcacaa TTGCCCTTTTTTTTATTAGAGAAAATTGTT atattagaat tatcaaaaaa TTGCCCTTTTTTTTATTAGAGAAAATTGTT atattagaat tatcaaaaaa TTG_CTTTTTTTTTTTAGAATAATTATGTT cagaggataa	(0) 32 bp (7)	MHMR*
P9-6	CIP 0.25 μM	<i>why1</i> <i>why3</i>	9948 Recombinant 30864	agaataatcc GTTTCGTTTTTTTACTTTC tctgaagta gttaagtttc GTTTCGTTTTTTTACTTTC tctgaagta gttaagtttc GTTTCGTTTTTTTACTTTC cttgatagat	(0) 20 bp (3)	MHMR*
P9-7a	CIP 0.25 μM	<i>why1</i> <i>why3</i>	9584 Recombinant 30885	ttttaacaat AGGAAAGTAAAAAAA gcgaaaagcg caatctatca AGGAAAGTAAAAAAA gcgaaaagcg caatctatca AGGAAAGTAAAAAAA aacgaaactt	(0) 16 bp (0)	MHMR
P9-7b	CIP 0.25 μM	<i>why1</i> <i>why3</i>	31015 Recombinant 30829	ttttctaaaa AAAAAAAA gcaaatTTTT ggtctaacat AAAAAAAA gcaaatTTTT ggtctaacat AAAAAAAA aaagaataaa	(0) 9 bp (0)	MHMR
P9-8	CIP 0.75 μM	Col-0	9472 Recombinant 31015	cacaattgcc CTTTTTTTTTTTATTAGAGAAAATT tgttatatta caaaaatttg CTTTTTTTTTTTATTAGAGAAAATT tgttatatta caaaaatttg CTTTTTTTTTTTATTAGAGAAAATT atgtttcaga	(0) 23 bp (3)	MHMR*
P9-9	CIP 0.75 μM	Col-0	9933 Recombinant 31046	tcaaaaatta AGATAATCCGTTTC gttttttata tattgttcag AGGATAATCCGTTTC gttttttata tattgttcag AGGATAATCCGTTTC tcttcatcac	(1) 15 bp (1)	MHMR*

Table 3-SI (continued)

P9-10	CIP 0.75 μ M	<i>why1</i> <i>why3</i>	9933 Recombinant 31046	tcaaaaatta AGATAATCCGTTTC gttttttata tattgttcag AGGATAATCCGTTTC gttttttata tattgttcag AGGATAATCCGTTTC tcttcacac	(1) 15 bp	MHMR*
P9-11	CIP 0.75 μ M	<i>why1</i> <i>why3</i>	9948 Recombinant 30859	agaataatcc GTTTCGTTTTTTT--ATACITTC tctgaagta agagggttaa GTTTCGTTTTTTTACTTTC tctgaagta agagggttaa GTTTCGTTTTTTTACTTTC cttgatagat	(3) 22 bp	MHMR*
P9-12	CIP 0.75 μ M	<i>why1</i> <i>why3</i>	9883 Recombinant 30757	ctagaatttc TTAAAAATTAATTAATAA ttaagagta aatattaaaa TTAATAATAATAATAA aaaaactaaa	(2) 18 bp	MHMR*
Plasmid reaction 10 with primers 60182REV and 50251FOR¹ – Deletion Events						
P10-1	CIP 0 μ M	<i>why1</i> <i>why3</i>	59060 Recombinant 50621	gctttttttg TACCTAGTCTTCTTATTTCTATTTCTATT cggttctatt atctgattaa TAACTAGTAATTTTTTTTATTTCTGTTTCTATT cggttctatt atctgattaa TAACTAGTAATTTTTTTTATTTCTGTTTCTATT taaaaaatt	(7) 33 bp	MHMR*
P10-2	CIP 0 μ M	<i>why1</i> <i>why3</i>	59724 Recombinant 51123	gagttacttt GAAACAAAAAAGGAAA gtcattttt tagaataaaa GAATACAAAAAAGGAAA gtcattttt tagaataaaa GAATACAAAAAAGTAAA aaaaattatc	(1) 19 bp	MHMR*
P10-3	CIP 0 μ M	<i>why1</i> <i>why3</i>	59726 Recombinant 50847	ctttgaaac AAAAAAAGGAAA gtcattttt gggatacaaa AAAAAAAGGAAA gtcattttt gggatacaaa AAAAAAAGGAAA attatgaact	(0) 15 bp	MHMR*
P10-4	CIP 0 μ M	<i>why1</i> <i>why3</i>	59953 Recombinant 50474	cgtccagca AATCATTTTTTTC cacaaggat aggttaatcg AATCATTTTTTTC cacaaggat aggttaatcg AATCATTTTTTTC gtttggtgt	(0) 14 bp	MHMR
P10-5	CIP 0.25 μ M	<i>why1</i> <i>why3</i>	59724 Recombinant 51123	gagttacttt GAAACAAAAAAGGAAA gtcattttt tagaataaaa GAATACAAAAAAGGAAA gtcattttt tagaataaaa GAATACAAAAAAGTAAA aaaaattatc	(2) 19 bp	MHMR*
P10-6	CIP 0.25 μ M	<i>why1</i> <i>why3</i>	59726 Recombinant 50847	ctttgaaac AAAAAAAGGAAA gtcattttt gggatacaaa AAAAAAAGGAAA gtcattttt gggatacaaa AAAAAAAGGAAA attatgaact	(0) 15 bp	MHMR*
P10-7	CIP 0.25 μ M	<i>why1</i> <i>why3</i>	59726 Recombinant 50845	ctttgaaac AAAAAAAGGAAA gtcattttt tggggataca AAAAAAAGGAAA gtcattttt tggggataca AAAAAAAGGAAA attatgaact	(0) 16 bp	MHMR*
P10-8	CIP 0.25 μ M	<i>why1</i> <i>why3</i>	60097 Recombinant 50836	tattttcgt TGGGGAT ttcctggaaa caagattgag TGGGGAT ttcctggaaa caagattgag TGGGGAT acaaaaaaa	(0) 7 bp	MHMR
P10-9	CIP 0.25 μ M	<i>why1</i> <i>why3</i>	59953 Recombinant 50474	cgtccagca AATCATTTTTTTC cacaaggat aggttaatcg AATCATTTTTTTC cacaaggat aggttaatcg AATCATTTTTTTC gtttggtgt	(0) 14 bp	MHMR
P10-10	CIP 0.75 μ M	Col-0	59846 Recombinant 51200	aattctcgt GGC ctttacccta ttcttagta GGC ctttacccta ttcttagta GGC tatacggact	4 bp	NHEJ- like
P10-11	CIP 0.75 μ M	Col-0	59953 Recombinant 50474	cgtccagca AATCATTTTTTTC cacaaggat aggttaatcg AATCATTTTTTTC cacaaggat aggttaatcg AATCATTTTTTTC gtttggtgt	(0) 14 bp	MHMR
P10-12	CIP 0.75 μ M	<i>why1</i> <i>why3</i>	58619 Recombinant 50471	catttttctt TC ² AATCATTTTTTTC taccgatatt ttcagggttaa TCGAATCATTTTTTTC taccgatatt ttcagggttaa TCGAATCATTTTTTTC cgttggtgt	(1) 16 bp	MHMR*
P10-13	CIP 0.75 μ M	<i>why1</i> <i>why3</i>	59398 Recombinant 50631	atcaactttt TTTTTT attttaaaaa taactagtaa TTTTTT attttaaaaa taactagtaa TTTTTT tattctgtt	(0) 7 bp	MHMR

Table 3-SI (continued)

P10-14	CIP 0.75 μ M	KO1/3	59719 Recombinant 50843	tactttgaaa ACAAAAAAAAAA ggaagtcac gagtggggat ACAAAAAAAAAA ggaagtcac gagtggggat ACAAAAAAAAAA aaagaagta	(0) 10 bp (0)	MHMR
P10-15	CIP 0.75 μ M	KO1/3	59953 Recombinant 50474	cgtccagca AATCATTTTTTTTC cacaagggat aggtaatcg AATCATTTTTTTTC cacaagggat aggtaatcg AATCATTTTTTTTC gtttggtgt	(0) 14 bp (0)	MHMR
P10-16	NOV 0 μ M	KO1/3	59724 Recombinant 51123	gagttacttt GAAACAAAAAAAAAGGAAA gtcataat tagaataaa GAATACAAAAAAAAAGGAAA gtcataat tagaataaa GAATACAAAAAAAAAGGAAA aaattatc	(1) 19 bp (1)	MHMR*
P10-17	NOV 0 μ M	KO1/3	59726 Recombinant 50847	ctttgaaac AAAAAAAAAAGGAAAGT catatttct ggggatacaa AAAAAAAAAAGGAAAGT catatttct ggggatacaa AAAAAAAAAAGGAAAGT attatgaact	(0) 15 bp (1)	MHMR*
P10-18	NOV 0 μ M	KO1/3	59953 Recombinant 50474	cgtccagca AATCATTTTTTTTC cacaagggat aggtaatcg AATCATTTTTTTTC cacaagggat aggtaatcg AATCATTTTTTTTC gtttggtgt	(0) 14 bp (0)	MHMR
P10-19	NOV 100 μ M	KO1/3	59724 Recombinant 51123	gagttacttt GAAACAAAAAAAAAGGAAA gtcataat tagaataaa GAATACAAAAAAAAAGGAAA gtcataat tagaataaa GAATACAAAAAAAAAGGAAA aaattatc	(1) 19 bp (1)	MHMR*
P10-20	NOV 100 μ M	KO1/3	59953 Recombinant 50474	cgtccagca AATCATTTTTTTTC cacaagggat aggtaatcg AATCATTTTTTTTC cacaagggat aggtaatcg AATCATTTTTTTTC gtttggtgt	(0) 14 bp (0)	MHMR
Plastid reaction 11 with primers 31381REV and 19309FOR						
Plastid reaction 12 with primers 69633REV and 79920FOR						
Plastid reaction 13 with primers 50513REV and 68821FOR						
P13-1	CIP 0 μ M	Col-0	50427 Recombinant 69421	agtcaattcc AATCGAGTGGATTGG gcaagggata ttttctaaaa AATCGAGTGGATTGG gcaagggata ttttctaaaa AAT_GGAGTTGATTGG atttgcacca	(1) 16 bp (1)	MHMR*
P13-2	CIP 0 μ M	KO1/3	50427 Recombinant 69421	agtcaattcc AATCGAGTGGATTGG gcaagggata ttttctaaaa AATCGAGTGGATTGG gcaagggata ttttctaaaa AAT_GGAGTTGATTGG atttgcacca	(1) 16 bp (1)	MHMR*
P13-3	CIP 0 μ M	KO1/3	50394 Recombinant 69096	aatgtggaac AAATA cccgccttagt agaattggat AAATA cccgccttagt agaattggat AAATA gtgggttcca	(0) 5 bp (0)	MHMR
P13-4	CIP 0.75 μ M	Col-0	50427 Recombinant 69421	agtcaattcc AATCGAGTGGATTGG gcaagggata ttttctaaaa AATCGAGTGGATTGG gcaagggata ttttctaaaa AAT_GGAGTTGATTGG atttgcacca	(1) 16 bp (1)	MHMR*
P13-5	CIP 0.75 μ M	KO1/3	50427 Recombinant 69421	agtcaattcc AATCGAGTGGATTGG gcaagggata ttttctaaaa AATCGAGTGGATTGG gcaagggata ttttctaaaa AAT_GGAGTTGATTGG atttgcacca	(0) 16 bp (2)	MHMR*

^a Treatment of plants with ciprofloxacin (CIP) or novobiocin (NOV) at the indicated concentrations.

^b The junction endpoint is the position of the first nucleotide preceding the DNA junction. The accession numbers for the mitochondrial and plastid genomes are NC_001284 and NC_000932, respectively.

^c Alignment of the sequenced DNA junctions (center) with homologous sequences (top and bottom). Mismatches are underlined.

^d Product types were classified according to junction length with NHE-J-like for junction of less than five nucleotides, MHMR for junctions of five or more nucleotides and MHMR* for junctions of five or more nucleotides containing at least one mismatch.

Table 3-SII. Related DNA rearrangements detected in the mitochondrial and plastid genomes

Same event in different genotypes or in different treatments
M7-1, M7-10
M11-1, M11-3
M12-1, M12-2, M12-4, M12-5
M12-3, M12-6
M16-1, M16-4
M18-2, M18-6
M18-1, M18-5
M20-1, M20-9
P1-2, P1-3, P1-4, P1-5, P1-6
P2-1, P2-4, P2-12, P2-20
P2-3, P2-6, P2-16, P2-22, P2-25
P2-5, P2-15, P2-23
P2-10, P2-21
P3-1, P3-6
P3-2, P3-8
P4-2, P4-5
P5-1, P5-2, P5-4
P6-1, P6-5, P6-9, P6-13
P6-2, P6-7, P6-11
P6-4, P6-8, P6-10, P6-12, P6-14
P7-1, P7-2, P7-3, P7-5
P8-2, P8-3, P8-5, P8-7, P8-9, P8-10, P8-12
P8-4, P8-6
P9-2, P9-4, P9-11
P9-9, P9-10
P10-2, P10-16, P10-19
P10-3, P10-6, P10-17
P10-4, P10-9, P10-11, P10-15, P10-18, P10-20
P13-1, P13-2, P13,4
Same event but with different mismatches
M3-6 vs M3-7
P9-6 vs (P9-2, P9-4 and P9-11)
(P10-2, P10-16 and P10-19) vs P10-5
P13-5 vs (P13-1, P13-2 and P13-4)
Events with at least one nearby breakpoint (less than 10 nts)
M15-1, M15-3, M15-4, M15-6
M20-7, M20-8
P9-1 vs (P9-2, P9-4 and P9-11)
P9-5 vs P9-8
(P10-4, P10-9, P10-11, P10-15, P10-18 and P10-20) vs P10-12
P10-8 vs P10-14

Table 3-SIII. DNA rearrangement statistics

Treatment	Type of DNA Rearrangement	Plastid Events ^a		Mitochondrial Events ^b	
		Col-0	<i>why1 why3</i>	Col-0	<i>why2</i>
CIP 0 μ M	NHEJ-like	0	0	0	0
	MHMR	2	18	0	0
	Total Events	2	18	0	0
CIP 0.25 μ M	NHEJ-like	1	0	0	1
	MHMR	7	26	1	13
	Total Events	8	26	1	14
CIP 0.75 μ M	NHEJ-like	7	4	11	8
	MHMR	10	26	10	25
	Total Events	17	30	21	33
NOV 0 μ M	NHEJ-like	0	1	0	0
	MHMR	0	11	0	0
	Total Events	0	12	0	0
NOV 100 μ M	NHEJ-like	0	0	0	0
	MHMR	1	8	0	0
	Total Events	1	8	0	0

NHEJ-like are DNA junctions with four or less homologous base pairs. MHMR are DNA junctions with five or more homologous base pair. CIP relates to plant treated with ciprofloxacin whereas NOV relates to plants treated with novobiocin. DNA rearrangements containing multiple DNA junctions were counted as multiple events.

^a Plastid events are based on 13 primer pairs for ciprofloxacin treatment and 10 primer pairs for novobiocin treatment.

^b Mitochondrial events are based on 20 primer pairs for ciprofloxacin treatment and 10 primer pairs for novobiocin treatment.

Table 3-SIV. DNA rearrangement products detected in the mitochondrial and plastid genomes and used as probes in DNA gel blots.

Plant treatment ^a	Plant Genotype	Junction Endpoints ^b	Junction Sequences ^c	Length of Junctions (Mismatches)	Junction Types ^d	Sizes of Hybridization Targets ^e
astid reaction A with primers 4434REV and 15643FOR						
CIP	KO1/3	4108	cccaaatgg AATTCTTT-TTTTTTTTC gacatcaaaa	(1)		4594 nt
0.75 μ M	Recombinant	16052	aataaaagcta AATTCTTTTTTTTTTC gacatcaaaa	19 bp (0)	MHMR*	913 nt (R) 6739 nt
astid reaction B with primers 49386REV and 68817FOR						
CIP	KO1/3	49165	tttttatttc CCCAC acctttttta	(0)		5872 nt
0.75 μ M	Recombinant	69006	ctctaaccct CCCAC acctttttta	6 bp (0)	MHMR	514 nt (R) 4474 nt
itochondrial reaction A with primers 125781REV and 180454FOR						
CIP	KO2	125068	acttacctag CGGAAGAAAGGG ctaggaccta	(0)		3887 nt
0.75 μ M	Recombinant	180866	tggagactga CGGAAGAAAGGG ctaggaccta	13 bp (0)		4862 or 5155 nt (R)
		46701	tggagactga CGGAAGAAAGGG gtctctccct	(0)	MHMR	2345 nt 2638 nt
itochondrial reaction B with primers 187018REV and 349054FOR						
CIP	KO2	335715	cgttgattg ACTTCACATAGCTTTTCCTCTGATAGCTGGAAG	(0)		3764 nt
0.75 μ M	Recombinant	349869	ttcttctttt ACTTCACATAGCTTTTCCTCTGATAGCTGGAAG	37 bp (0)	MHMR	2914 nt (R) 6566 nt

^a Treatment of plants with ciprofloxacin (CIP) or novobiocin (NOV) at the indicated concentrations.

^b The junction endpoint is the position of the first nucleotide preceding the DNA junction. The accession numbers for the mitochondrial and plastid genomes are NC_001284 and NC_000932, respectively.

^c Alignment of the sequenced DNA junctions (center) with homologous sequences (top and bottom). Mismatches are underlined.

^d Product types were classified according to junction length with NHE-J-like for junction of less than five nucleotides, MHMR for junctions of five or more nucleotides and MHMR* for junctions of five or more nucleotides containing at least one mismatch.

^e Sizes of HindIII-digested DNA molecules that are expected to hybridize with the PCR product. R indicates the recombinant form.

Table 3-SV. Oligonucleotides used in this study

Reaction	Oligonucleotide Name	Oligonucleotide Sequence
Mitochondrial PCR primers		
1	20041REV	GCG CAT ACC TTA TCC GAT CGT
	30001FOR	CTT CCG CTA TTG GTG GAC TGT
2	30001REV	ACA GTC CAC CAA TAG CGG AAG
	50341FOR	GAG CTA GAT AGC ACG GCT TTG
3	55021REV	AAG GCT AGT CGT CAA CGA GCA
	65041FOR	GGA TCT GAG CTG TAG AAA CCC
4	130413REV	GCA CAG TCG AGT GCT TTA CCT
	151514FOR	CTG GAT CTC GCT CAT GTC TCT
5	180454REV	GTT TAC ATG GGG CTG GTA GGT
	200500FOR	CCC ACT GAG TTG GTA AGC AAG
6	235252REV	AAA CAG TGC GGT GAG CTC TAG
	245966FOR	GCA AGA TGA TCA GTC CGA GAG
7	305346REV	GTT TCG CAG CAC GGA GTA TAC
	315513FOR	CAT CTG AGG CAC CTC TAT TGG
8	165904REV	TGT ACT GGC TTC CGC ACT ACT
	200500FOR	CCC ACT GAG TTG GTA AGC AAG
9	155034REV	CTG ATG GTG TTC GTG GTA GGT
	200500FOR	CCC ACT GAG TTG GTA AGC AAG
10	10081REV	CAT GTC GAA GGA ACT CGG CAA
	15001FOR	CCG GAA GGT ATA GTG TGA AGG
11	20461REV	GCG CAT ACC TTA TCC GAT CGT
	24961FOR	TGA TTC CGT CTG ACT CCT CCT
12	30001REV	ACA GTC CAC CAA TAG CGG AAG
	80161FOR	ACG TGC AAG TTT CCC TGC ATG
13	84901REV	CAA AAC ACT CCA GAT CCG CAC
	151514FOR	CTG GAT CTC GCT CAT GTC TCT
14	210239REV	GGA AGG CTT TGT CTA CGG TCT
	245966FOR	GCA AGA TGA TCA GTC CGA GAG
15	1REV	GGC TAA CCT GTT TCG AAC GGA
	4981FOR	CCA GGA AAG CCT AGC TAC TTC
16	125781REV	CCA GAC TCG GAT GCT AGA GAA
	115816FOR	TTA CCT ACC CGG CTA CTG ACA
17	135180REV	ATC TCT AAG TGC TCG CAG CAC
	140053FOR	TTC CCT ATT CCT CGC CTA CTC
18	171214REV	CAT TCT AGC CCG AGA GGA ACT
	180454FOR	ACC TAC CAG CCC CAT GTA AAC
19	245966REV	CTC TCG GAC TGA TCA TCT TGC
	239584FOR	AGG GCT ACC TTT CTT CGA CTG
20	270785REV	ACA CAA ACT GAC CGA GCT AGC
	275583FOR	GCA CAG TGG ATT CAT CTA GGG

Table 3-SV (continued)

<i>cox1</i>	cox1REV	GTA	ACG	TCC	GTT	CCG	TGA	TCT	
	cox1FOR	GCT	AGC	TCA	TGG	CAG	GAA	ATC	
A	125781REV	CCA	GAC	TCG	GAT	GCT	AGA	GAA	
	180454FOR	ACC	TAC	CAG	CCC	CAT	GTA	AAC	
B	187018REV	AAG	AAG	CTA	CCT	CCT	TCC	GTG	
	349054FOR	AAT	CCG	ATA	GCT	CTA	CCA	GCC	
Repeat A	AFOR	CAC	TGA	ACC	GAC	TTG	AAT	CTG	AAC
	AREV	GAA	AGC	CAT	TTG	CAC	CAG	GTT	TAT
Repeat D	DFOR	AGT	GAT	CTG	TTC	ATC	TAA	CTC	A
	DREV	TAC	TAC	TAC	CTC	GTC	CAT	TG	
Repeat F	FFOR	CAC	GAG	GAA	TGG	AAA	GAA	ACA	T
	FREV	GCG	CAC	AAA	CCA	CTC	TAA	AG	
Plastid PCR primers									
1	31138REV	AAT	AGC	CTA	GCG	CAT	CGA	ACG	
	19181FOR	CAG	CAT	CTG	ATG	TGC	GTA	CAG	
2	40130REV	GGG	AAG	AGG	AGG	AAC	ATG	TCA	
	50251FOR	GGC	CCC	TTC	CTA	ATT	GGA	GAT	
3	84761REV	CCT	TAG	GCA	CGG	CCA	TAC	ATA	
	95260FOR	AAC	CTG	CTA	GTG	GAG	GAA	GAC	
4	105961REV	GCG	TTA	CTC	AAG	CCG	ACA	TTC	
	116007FOR	AAA	CCA	CTC	ATG	CCC	GGC	AAT	
5	69633REV	GAG	GAA	CAT	GTA	CTC	GGG	TGT	
	59291FOR	AGG	CCT	AGT	CTT	TCC	GGG	AAT	
6	111330REV	TCA	ACA	GCG	GGA	TTA	ACC	GCA	
	121141FOR	AAC	GTC	ATC	CGG	TTA	CCG	GTT	
7	50513REV	GAG	ACG	TCT	ACC	ACA	ACA	ACC	
	59311FOR	AGG	CCT	AGT	CTT	TCC	GGC	AAT	
8	68821REV	ACG	CGG	ATA	TAT	AGC	TCG	GAG	
	79920FOR	CCA	TGC	CGT	TCG	GAA	GTG	ATT	
9	10230REV	TGG	CAA	GAG	GTC	AAC	GAT	TGC	
	30511FOR	GCT	GCC	TCC	TTG	AAA	GAG	AGA	
10	60182REV	ATA	TAA	AGG	ACA	CGC	CGG	GCA	
	50251FOR	GGC	CCC	TTC	CTA	ATT	GGA	GAT	
11	31381REV	CCG	ATG	ACT	TAC	GCC	TTA	CCA	
	19309FOR	TCG	TCT	GCT	AAG	ACA	CGA	CCA	
12	69633REV	GAG	GAA	CAT	GTA	CTC	GGG	TGT	
	79920FOR	CCA	TGC	CGT	TCG	GAA	GTG	ATT	
13	50513REV	GAG	ACG	TCT	ACC	ACA	ACA	ACC	
	68821FOR	CTC	CGA	GCT	ATA	TAT	CCG	CGT	
A	4434REV	TTT	ATC	CCT	GAC	GCA	CAC	ACG	
	15643FOR	CCA	TTT	CCT	CGT	ACC	ATG	TCC	
B	49386REV	TGT	ACG	CAG	TCC	GCA	TAT	TGG	
	68817FOR	GTT	TTG	TTC	TAC	GCC	TCC	GAG	
<i>ycf2</i>	ycf2REV	TGT	TTC	GCC	TCT	TAC	TCG	GAG	
	ycf2FOR	GAT	CTC	TGA	GAG	CTG	TTT	CCG	

CHAPITRE 4

**Une lysine des protéines Whirly de plantes est nécessaire
pour la protection contre les dommages à l'ADN ainsi
que pour la formation de superstructures**

Article 3 : A Conserved Lysine Residue of Plant Whirly Proteins is Necessary for Higher
Order Protein Assembly and Protection Against DNA Damage

Article en préparation

A Conserved Lysine Residue of Plant Whirly Proteins is Necessary for Higher Order Protein Assembly and Protection Against DNA Damage

Laurent Cappadocia, Jean-Sébastien Parent, Jurgen Sygusch, and Normand Brisson

Department of Biochemistry
Université de Montréal
PO Box 6128, Station Centre-Ville
Montréal, Québec
H3C 3J7, Canada

Article en préparation

4.1 Mise en contexte

En analysant l’empaquetage des protéines WHY2 dans la maille cristalline, nous nous sommes aperçus que celles-ci ne sont pas seulement agencées en tétramères, mais aussi en hexamères de tétramères (ou 24-mères). Nous nous sommes alors demandé si ces entités existaient en dehors de l’état cristallin. L’utilisation de la technique de microscopie à force atomique a confirmé que WHY2 peut s’agencer en 24-mère *in vitro* lors de sa liaison à de longues séquences d’ADN monocaténaire. Dans les cristaux, l’assemblage des protéines Whirly en 24-mères est stabilisé par des interactions dépendantes d’un résidu lysine. La substitution de cet acide aminé par une alanine n’affecte ni la structure de la protéine ni sa liaison à de courtes séquences d’ADN. Cependant, la protéine mutée ne forme plus de 24-mères *in vitro*. Pour comprendre le rôle de cet assemblage *in vivo*, nous avons effectué des expériences de complémentation chez *Arabidopsis*. Nous avons réintroduit la protéine WHY1 de type sauvage ou la protéine WHY1 K91A chez des plantes n’exprimant ni WHY1 ni WHY3 et nous n’avons observé un phénotype de complémentation que dans le premier cas. Ceci suggère que l’assemblage en 24-mères est important *in vivo* pour la réparation de l’ADN.

Les résultats contenus dans ce chapitre seront bientôt soumis pour publication. J’ai effectué les expériences de radiocristallographie et les analyses par microscopie à force atomique. J’ai également travaillé en étroite collaboration avec Jean-Sébastien Parent pour la production et la caractérisation phénotypique et moléculaire des plantes d’*Arabidopsis* transgéniques exprimant le gène WHY1 ou le gène WHY1 K91A. J’ai rédigé le présent manuscrit qui a par la suite été corrigé par tous les auteurs.

4.2 Abstract

All organisms have evolved specialized DNA repair mechanisms in order to protect their genome against detrimental lesions such as DNA double-strand breaks. In plant organelles, these damages are repaired either through recombination or through a microhomology-mediated break-induced replication pathway. Whirly proteins have been shown to protect both chloroplast and mitochondria genomes against illegitimate repairs.

In the current view, tetrameric Whirly proteins are proposed to bind single-stranded DNA and to prevent spurious annealing of resected DNA molecules with other regions in the genome. In this study, we add a new layer of complexity to this model by showing through atomic force microscopy that tetramers of the potato Whirly protein WHY2 can further assemble into hexamers of tetramers, or 24-mers, upon binding long DNA molecules. The hexamerization of Whirly tetramers is dependent on tetramer-tetramer interactions mediated by K67, a lysine residue strongly conserved among Whirly proteins. Indeed, the mutation of this residue abolishes the formation of 24-mers. The crystal structures of the WHY2 K67A variant both in the free form and bound to a dT₃₂ oligonucleotide demonstrate that the mutated protein is structurally similar to the wild-type protein and that binding to short DNA molecules is not affected by the K67A mutation. Furthermore, we show that an *Arabidopsis* Whirly protein mutated for this lysine is unable to rescue the sensitivity of a Whirly-less mutant plant to a DNA double-strand break inducing agent.

4.3 Introduction

Plants must protect the integrity of genomic DNA located in three distinct compartments (the nucleus, mitochondria and plastids) against deleterious DNA lesions such as DNA double-strand breaks (DSBs). Although nuclear DNA repair mechanisms have been the subject of intense studies (1,2), very little is known about plant organelle DNA repair pathways (2-4). These pathways must however be well developed since both mitochondria and plastid DNA are routinely exposed to high levels of DNA damaging reactive oxygen species resulting from the intense metabolic activity that takes place in these compartments (5,6).

Until very recently, homologous recombination was the only mechanism proposed to be involved in the repair of DSBs in organelles. Indeed, early studies postulated the existence of a DNA recombination machinery (7,8) based on biochemical evidences such as the presence of functional DNA recombinases in both plastid and mitochondria (9-11), the detection of strand exchange activity in purified organelles (12,13) and the monitoring of recombination events following DNA damage (14,15). DNA repair through recombination is expected to be a very efficient process in both plant mitochondria and plastids as the

genomes of these organelles are highly polyploid, thus providing intact DNA molecules that serve as template for the repair of the damaged DNA (16). Recently, we demonstrated that treatment of *Arabidopsis thaliana* plants with ciprofloxacin, a small molecule that produces DSBs in organelles, resulted in the formation of DNA rearrangements in both mitochondria and plastids (15). These DNA rearrangements showed microhomology at their breakpoint junctions, suggesting that they are the result of DNA repair through a microhomology-mediated break-induced replication (MMBIR) pathway (15), which is also active in bacteria, yeast and humans (17).

Most of the players involved in the MMBIR pathway have yet to be identified. We recently demonstrated that Whirly proteins protect against MMBIR events in both chloroplasts and mitochondria (15). Whirlies constitute a small family of single-stranded DNA binding proteins (SSB) that are mainly found in plants (18). They assemble into tetramers resembling whirligigs (19) and perform numerous cellular functions, including gene regulation, telomere maintenance, RNA editing and DNA repair (15,18,20-22). All plants examined to date possess at least two nuclear genes coding for Whirly proteins. The products of these genes are directed to the mitochondria and plastids (23). In *Arabidopsis*, where there are three genes, WHY1 and WHY3 are directed to the chloroplast whereas WHY2 is directed to the mitochondria. In absence of both WHY1 and WHY3, DNA rearrangements mediated by microhomologies occur spontaneously in the chloroplasts and often correlate to development of white-yellow variegated leaves (22). Induction of DSBs by treatment with the gyrase inhibitor ciprofloxacin leads to a further increase in microhomology-mediated DNA rearrangements in the chloroplasts of *why1 why3* mutant plants that lack both *WHY1* and *WHY3* genes (15). Contrasting with the situation in *why1 why3* plants, the inactivation of *WHY2* does not result in any visible or molecular phenotype in absence of DNA stress (24). Upon induction of DSBs, however, plants lacking *WHY2* accumulate microhomology-mediated mitochondrial DNA rearrangements more readily than wild type *Arabidopsis* plants (15). The different behavior between *why2* and *why1 why3* plants could be due to the partial functional redundancy provided by other SSBs present in plant mitochondria (15,25). Whirly proteins are believed to modulate the MMBIR pathway either by binding the resected DNA that forms during processing of DSBs, thus favoring error-free repair through homologous recombination (15), or by binding ssDNA

from all over the genome (21,22,24) and thus preventing spurious annealing of DNA through microhomologous sequences (15).

In contrast to most SSBs that use an OB fold as their main ssDNA binding platform (26), Whirly proteins use a Whirly domain (18,19). This domain is typically composed of two four-stranded β -sheets and two α helices. These helices constitute the core of the protein against which stacks the β -sheets in a whirligig-like manner (19). Determination of the atomic structure of the potato Whirly protein WHY2 in complex with different ssDNA molecules shed light on the mechanism of DNA binding (15). Indeed, WHY2 binds ssDNA by establishing hydrogen bonds with the phosphate moiety of the ssDNA and by stacking nucleobases in between the β -sheets of adjacent subunits. This enables binding of ssDNA with high affinity and little sequence specificity as the bases are not brought in close contact with the protein surface. SSBs can bind DNA with various degrees of cooperativity (27,28). For bacterial SSB, the cooperative binding to long ssDNA molecules is due to the formation of a protein filament (28,29). In the case of Whirly proteins the formation of a superstructure upon binding long DNA molecules has not yet been investigated.

A KGKAAL sequence motif is a hallmark of Whirly proteins as this motif is perfectly conserved among plant Whirlies (18,19). This motif is important for ssDNA binding as its replacement with an alternate sequence abolishes the interaction with ssDNA (19). Rather than interacting directly with DNA, many residues within this motif seem important to maintain the stability of the β -sheet that contains numerous DNA-interacting residues. The role of the second lysine of the KGKAAL motif is however enigmatic. Indeed, in the crystal structure of the potato WHY1 protein (PDB 1L3A), the ϵ amino group of this conserved residue points in the general direction of the solvent in three out of four subunits and does not appear to contribute to the overall stability of the tetramer.

In this manuscript, we show that WHY2, a *Solanum tuberosum* Whirly protein located in the mitochondria, binds long ssDNA in a cooperative manner through hexamerization of Whirly tetramers. This results in the formation of hollow protein shells of 12 nm diameter with 432 symmetry. We also highlight the unique contribution of the second lysine residue

of the KGKAAL motif to the formation of 24-mers. Mutation of the equivalent lysine in At-WHY1 partially abolishes the complementation of the etiolated/variegated phenotype in *why1 why3 Arabidopsis* plants treated with ciprofloxacin, highlighting the functional importance of this residue *in planta*.

4.4 Experimental procedures

4.4.1 Preparation of WHY2 and WHY2 K67A

Details concerning cloning, expression and purification of WHY2 are reported elsewhere (30). The K67A mutant of WHY2 was generated using the method developed by Zheng (31). The mutation was confirmed by DNA sequencing. Expression and purification were performed as for the non-mutated protein.

4.4.2 Data collection and structure determination

WHY2 K67A in the free form (forms I and II; 20 mg/ml) was crystallized by the hanging drop vapor diffusion method using a precipitant solution containing 100 mM MOPS pH 7.0, 18-24 % (w/v) PEG1000 and 100 mM $\text{NH}_4\text{H}_2\text{PO}_4$. The WHY2 K67A-dT₃₂ (10 mg/ml) complex was crystallized by the hanging drop vapor diffusion method using a precipitant solution containing 100 mM Tris/HCl pH 8.0, 15 % (w/v) PEG6000 and 1.2 M LiCl. Diffraction data were collected using an ADSC Quantum 315 detector at beamline X29 at the National Synchrotron Light Source (NSLS) at the Brookhaven National Laboratory (BNL, USA) or using a Rayonix MX300 detector at beamline 08ID-1 at the Canadian Light Source (CLS). Initial phases for WHY2 in the free form and the WHY2 K67A-dT₃₂ complex were obtained by molecular replacement using WHY2 (PDB 3N1H) and WHY2-dT₃₂ (PDB 3N1I) as a search template. Phases were improved by iterative cycles of model building with Coot (32) and refinement with CNS (33,34) and Phenix (35). Clear density was visible for protein residues 55-215 except for WHY2 K67A in the free form II. In this structure, clear density was visible for residues 54-141 and for residues 150-214. The residues 142 to 149 were omitted as they are disordered. Test data sets were randomly selected from the observed reflections prior to refinement. The figures were prepared with PyMOL (36).

4.4.3 Electrophoretic mobility shift assays

The apparent dissociation constant for WHY2 K67A proteins bound to DNA was determined by EMSA as described previously (15). The autoradiograms were scanned and the intensity of the bands corresponding to bound and free radiolabeled ssDNA was quantified using ImageJ (NIH, <http://rsb.info.nih.gov/ij/>). The data was plotted and analyzed using Prism 5 Demo (GraphPad Software). The apparent dissociation constants were calculated by fitting the plot of the fraction of ssDNA bound versus protein concentration to the Hill equation.

4.4.4 Agarose-based electrophoretic mobility shift assays

Binding reactions including 600 ng of M13mp18 ssDNA (USB) and the serially diluted WHY2 or WHY2 K67A protein were incubated for 20 min at 20°C in a buffer containing 20 mM Tris-HCl pH 8.0 and 20 mM NaCl. Reactions were run on a 0.7 % (w/v) agarose gel with 40 mM Tris, 20 mM Acetic acid, 1 mM EDTA pH 8.0 and 0.005 % (v/v) ethidium bromide at 20°C under 80 V constant voltage for one hour. DNA and protein-DNA complexes were visualized by UV transillumination.

4.4.5 Atomic force microscopy

AFM was performed in tapping mode in air using a JEOL JSPM-5200 microscope. Images of 512×512 pixels with a scan size of 1.5×1.5 μm were acquired at scan frequency of 1–1.5 Hz. Topographic and phase images were acquired simultaneously. Topographic images were used for height-measurement whereas phase images were used to identify features on the sample surface that were not easily distinguished by height (37). Raw AFM images were processed using a JEOL software package (WinSPM DPS software, JEOL Ltd.) and the Gwyddion software (Czech Metrology Institute, <http://gwyddion.net/>).

Imaging of the protein-DNA complexes was performed as described (38) using a 5- μl droplet of a solution containing 1 $\mu\text{g.ml}^{-1}$ WHY2 or WHY2 K67A, 20 mM Tris-HCl pH 8.0, 20 mM NaCl and 50 μM Spermidine. Imaging of the free protein was performed using a 5- μl droplet of a solution containing 1 $\mu\text{g.ml}^{-1}$ WHY2, 20 mM Tris-HCl pH 8.0,

20 mM NaCl and 100 μ M NiCl. All samples were deposited onto the surface of freshly cleaved mica for 1 min. The surface was then rinsed with a 0.02% uranyl acetate solution. The samples were then rapidly rinsed with pure water (Millipore) to obtain a clean surface after drying under a stream of nitrogen.

4.4.6 Plant material and growth conditions.

The *why1 why3* mutant line was described previously (15,22). To complement *why1 why3*, the entire coding sequence of WHY1, or its K91A mutant, was amplified by RT-PCR using total RNA from *Arabidopsis thaliana* ecotype Col-0 and cloned into the pGreen vector containing the BAR cassette and under the control of the 35S viral promoter (39). These constructs were individually cotransformed with the pSOUP vector into a GV3101 pMP90 *Agrobacterium tumefaciens* strain. *Arabidopsis why1 why3* plants were then transformed by the floral dip method as described (40). Seeds were harvested, planted on soil and screened for BASTA resistance. Two independent homozygous lines were obtained for each construct. The *why1 why3* #1 and the *why1 why3* #2 lines were derived from *why1 why3* 35S:WHY1 #1 and *why1 why3* 35S:WHY1 #2 following segregation of the transgene. For the complementation analysis, sterilized seeds of Col-0 (wild type), *why1 why3* and complemented lines were grown for 10 days on Murashige and Skoog basal medium (Sigma-Aldrich) containing 0.125 μ M ciprofloxacin. One-way ANOVA and Tukey's HSD test were used for multiple phenotype comparisons. Plants were analyzed by protein gel blot and electrophoretic mobility shift assay as described previously (22).

4.5 Results

4.5.1 WHY2 assembles into hexamers of tetramers in the crystals

In the crystal structure of the potato WHY2 protein (PDB 3N1H), K67 forms a hydrogen bond with the backbone carbonyl of the F138 residue of a symmetry-related tetramer (Figure 4-1A). Moreover, in the crystal structures of WHY2-DNA complexes (PDB 3N1I, 3N1J, 3N1K, 3N1L), K67 forms a hydrogen bond with the edge of a nucleobase of a symmetry-related tetramer (Figure 4-1B). Through these interactions, in all the WHY2

crystal structures, K67 contributes to the assembly of Whirly tetramers into hexamers of tetramers, or 24-mers, that are predicted to be stable in solution by the Protein Interfaces, Surfaces and Assemblies (PISA) service (41). In the 24-mer configuration, approximately half of the solvent-accessible surface in each WHY2 monomer is buried at the intersubunit contact. About 65% of the contact surfaces (2.2 nm^2) contribute to the tetramer formation whereas only 35% (1.1 nm^2) contribute to the further assembly of tetramers into 24-mers. This, plus the fact that the surface involved in the tetramer formation is more hydrophobic/aromatic than the one involved in the 24-mer assembly, suggests a difference in the stability of the two oligomeric states of WHY2.

The 24-mers form roughly spherical protein shells with 432 symmetry (Figure 4-1A and 4-1B) that measure 12 nm of diameter, have a 5 nm diameter hollow core and have both their N- and C-termini projecting outward from the sphere. Each protein shell has six 0.8 nm-wide pores that coincide with the four-fold axes. These pores allow access to an electropositive inner cavity (Figure 4-S1). Adjacent 24-mers interact with each other through their β -strands. In DNA-bound WHY2 complexes, binding of ssDNA takes place on the external surface of the spheres (Figure 4-1C).

We previously hypothesized that long ssDNA molecule could travel from one binding site to another within a single tetramer (15). Indeed, although WHY2 was crystallized with a 32 nucleotide long ssDNA molecules, each protein protomer only binds a nine nucleotide long fragment (15). Interestingly, the 5'- and 3'-ends of DNA fragments bound by adjacent 24-mers are separated by only 1.9 nm (Figure 4-1C). Three nucleotides can easily span this distance considering a mean of 0.6 nm between phosphorus atoms of neighboring nucleotides. Our present observation that DNA ends are located in close proximity between adjacent 24-mers thus raises the possibility that two pathways exist for the ssDNA in the crystals: one that goes to the adjacent protomer of the same tetramer and one that goes to an adjacent 24-mer (Figure 4-1C).

4.5.2 WHY2 assembles into hexamers of tetramers *in vitro* upon binding long DNA

To verify whether WHY2 assembles into 24-mers *in vitro*, we observed unbound and dT₃₂-bound WHY2 proteins using atomic force microscopy (AFM), a technique that allows estimation of the oligomeric state of particles through measurement of their heights. For WHY2 in the free form or in complex with a dT₃₂ oligonucleotide, only scattered particles with an average height of 3-5 nm were detected by AFM (Figure 4-2A and 4-2B). These particles are likely tetramers lying flat on the mica surface, consistent with a height of 4 nm for the WHY2 tetramer in the crystals and with the results of size-exclusion chromatography showing that WHY2 and WHY2-dT₃₂ complexes mainly exist as tetramers (Figure 4-S2). This indicates that, *in vitro*, WHY2 does not form 24-mers constitutively or upon binding small DNA molecules. In contrast, AFM revealed that complexes of WHY2 bound to M13mp18, a 7249 nucleotide long ssDNA molecule, existed as grouped particles (Figure 4-2C and 4-2D). These particles exhibited a bimodal distribution with average heights of 6.0 and 10.4 nm. A height of 6.0 ± 1.7 nm is similar to the heights measured for WHY2 (5.1 ± 2.0 nm) and WHY2-dT₃₂ (3.4 ± 1.4 nm) suggesting that these species could be DNA-bound tetramers while a height of 10.4 ± 1.9 nm is in agreement with that of the 24-mers observed in the crystal lattice of WHY2 (12 nm). This suggests that 24-mers of WHY2 may form *in vitro* either at very high local concentration (the concentration of Whirly proteins reaches 35 mM in the crystals) or upon binding long single-stranded DNA molecules.

4.5.3 K67 stabilizes the 24-mer assembly of WHY2

To assess the contribution of K67 to the stability of the 24-mers, we produced a K67A mutant of WHY2. This mutant assembles into tetramers both in the free form and upon binding of a dT₃₂ oligonucleotide (Figure 4-S2). Electrophoretic mobility shift assays (EMSA) experiments further reveal that this mutant binds dT₃₂ with an apparent dissociation constant of 3.6 ± 0.5 nM (Figure 4-3A). This is similar to the value observed for the binding of the nonmutated protein to dT₃₂ (4.5 ± 0.9 nM; (15)), thus suggesting that the K67A mutation does not affect the binding of individual tetramers to short ssDNA molecules. The binding of WHY2 K67A to M13mp18 ssDNA was then assessed by AFM (Figure 4-3B

and 4-3C). As for WHY2-M13mp18, the complexes existed as grouped particles. However, the average height of the particles of the WHY2 K67A-M13mp18 complex (2.9 ± 1.4 nm) was much lower than for WHY2-M13mp18 (10.4 ± 1.9 nm) and was consistent with that of DNA-bound tetramers lying flat on the mica surface (predicted size of 4 nm). This suggests that WHY2 K67A does not assemble into 24-mers upon binding long DNA.

To insure that the K67A mutation does not affect the overall structure of the protein, we solved the crystal structure of WHY2 K67A both in the free form (Figure 4-4A and Table 4-I) and bound to dT₃₂ (Figure 4-4B and Table 4-I). Consistent with the EMSA results, the dT₃₂-bound K67A mutant adopts the same conformation as the nonmutated protein (Figure 4-4B) with a root-mean square deviation (rmsd) of merely 0.2 Å for all aligned C α , explaining why the binding to dT₃₂ is not affected by this mutation. In the crystal structure of WHY2-dT₃₂ (PDB 3N1J), K67 stabilizes the F432 lattice by establishing a hydrogen bond with the edge of a nucleobase from a symmetry-related tetramer. Although these interactions are lacking in the crystal structure of the WHY2 K67A-dT₃₂ complex, other interactions involving the residues 78-84, 102-106, 112, 144-148, 152, 162-164, 173-182 still maintained the assembly of individual tetramers into 24-mers.

The unbound WHY2 K67A protein crystallized in two different space groups under the same chemical conditions: F432 and P4₂1₂. In these space groups, WHY2 K67A tetramers interact with symmetry-related tetramers through partially different sets of residues (residues 78-84, 102-106, 112, 144-148, 152, 162-164 and 173-182 in the P4₂1₂ space group and residues 53, 56, 83-84, 102-104, 106-112, 144, 160-161, 167, 173-182 in the F432 space group). These interactions promote the assembly of individual tetramers into 24-mers only in the F432 space group. Importantly, the proteins in both space groups are structurally similar to the nonmutated protein (Figure 4-4A), confirming that the K67A mutation does not perturb the overall structure of the protein. A small rearrangement can however be observed in the loop consisting of residues 139-145 in the free form structure of WHY2 K67A (form I) as compared to the nonmutated protein (Figure 4-4A). This rearrangement appears to stabilize the interaction between adjacent tetramers thereby partially compensating for the absence of a K67-F138 hydrogen bond in the F432 space group. Globally, these results suggest that the K67A mutation does not disturb the overall

structure of the protein, the binding to short DNA molecules or the capacity of WHY2 to assemble into 24-mer at very high protein concentration. However, this mutation is sufficient to prevent 24-mer formation at least *in vitro*.

4.5.4 Cooperative binding of WHY2 upon binding long DNA molecules

The close proximity of the DNA ends in the crystal structure of WHY2-DNA complexes along with the existence of interactions between adjacent 24-mers suggest that WHY2 could bind ssDNA in a cooperative manner. To verify this hypothesis, we performed an agarose-based EMSA experiment that enable qualitative evaluation of the cooperative binding of proteins to long DNA molecules (42). Increasing amounts of WHY2 was incubated with a fixed amount of M13mp18 ssDNA and the migration profile was analyzed on an agarose gel (Figure 4-5A). As the concentration of WHY2 was increased, DNA became trapped in a large protein-DNA complex that barely entered the gel (Figure 4-5A). This suggests that WHY2-ssDNA complexes form in an all-or-nothing manner characteristic of cooperative binding (42-44). In contrast, agarose-based EMSA experiments show that the K67A mutation abrogates the cooperative binding of WHY2 to ssDNA, as complexes of decreasing mobility were formed when increasing amounts of WHY2 K67A were added to M13mp18 (Figure 4-5B). Globally, these results indicate that the K67A mutation disrupts both the 24-mer assembly and the cooperative binding to long ssDNA.

4.5.5 Analysis of the K91A mutation in *Arabidopsis*

We next analyzed the effect of the lysine mutation *in vivo* in *Arabidopsis* as the biological function of Whirly proteins in DNA repair has mainly been studied in this organism. Precisely, we tested the ability of WHY1 and WHY1 K91A (the equivalent of K67A for At-WHY1) transgenes to complement the variegation/etiolation phenotype of *why1 why3* plants upon ciprofloxacin treatment. The *why1 why3* mutant background was chosen because, at low doses of ciprofloxacin, it develops extensive variegation/etiolation whereas the WT control is merely affected (15). Consistent with our previous results, upon treatment with 0.125 μ M ciprofloxacin, only one percent of WT plants developed variegated/etiolated leaves whereas, in the same conditions, 52 and 59 % of two *why1 why3*

lines developed variegation/etiolation (Figure 4-6A). Interestingly, two *why1 why3* lines complemented with *35S:WHY1* displayed partial rescue of the Whirly function with 18 and 10 % of leaves showing variegation/etiolation. By contrast, the leaves of two *why1 why3* lines genetically transformed with *35S:WHY1 K91A* displayed 38 and 43 % variegation/etiolation, respectively, which is more similar to the levels observed for the untransformed *why1 why3* lines (Figure 4-6A). Noteworthy, there are no significant differences ($p > 0.05$) between the two independent lines obtained for each mutant (see Table 4-SI for a complete statistical treatment of these data). There is however a significant difference ($p < 0.001$) between the variegation/etiolation percentage observed for *35S:WHY1* and the *35S:WHY1 K91A* line thus underlining the critical role of K91 *in vivo*. Figure 6B demonstrates that the WHY1 protein accumulates to similar levels in *35S:WHY1* and *35S:WHY1 K91A* lines, far above the endogenous level. Figure 6C indicates that the ssDNA binding activity is also similar in *35S:WHY1* and *35S:WHY1 K91A* lines and proportional to the protein level. Thus, consistent with our *in vitro* analyses of St-WHY2, no reduction in DNA binding activity was observed upon K91A mutation. Globally, these results suggest that the K91A mutation does not destabilize or reduce the DNA binding affinity of WHY1. Nevertheless, this mutation interferes with the biological function of Whirly proteins in DNA repair.

4.6 Discussion

We previously demonstrated that Whirly proteins protect *Arabidopsis* organelle DNA against the detrimental effects of ciprofloxacin by preventing DSB repair by the error-prone MMBIR mechanism (15). We now show that this protection is greatly reduced when the second lysine residue of the KGKAAL motif of At-WHY1 is mutated. This is at first glance intriguing since this mutation does not affect WHY1 stability nor its binding affinity to short ssDNA oligonucleotides. However, the structural analysis of the St-WHY2 protein provides clues to understand the possible function of this residue. Indeed, we established that the Whirly proteins are arranged in 24-mers in the crystals of WHY2 or WHY2-DNA complexes and that the second lysine residue of the KGKAAL motif contributes to the stabilization of this assembly. Furthermore, we confirmed that the 24-mer assembly which, *in vitro*, only forms upon binding long DNA sequences, is abolished when the K67 residue is mutated. Therefore, a possible role for this residue *in planta* could be to promote 24-

mer assembly and cooperative binding to DNA. These could be general properties of plant Whirly proteins as both the Whirly fold and the second lysine residue of the KGKAAL motif are strongly conserved among plant Whirlies (18,19).

Similarly to plant Whirlies, bacterial SSB proteins bind ssDNA with high positive cooperativity through the formation of a higher-order protein assembly (29,45). Indeed, in *Escherichia coli*, the SSB protein can exist in two states, (SSB)₃₅ and (SSB)₆₅, that bind ssDNA with different degrees of cooperativity (46). Several studies investigated the functional consequences of mutations in this protein and revealed a W54S mutation that stabilizes the (SSB)₃₅ state that binds ssDNA with high positive cooperativity (47,48). Interestingly, when the gene coding for this mutated protein is expressed in *E. coli*, the cells exhibit slow growth and increased UV sensitivity (47). The authors of this study concluded that DNA repair was more dependent on the poorly cooperative (SSB)₆₅ state whereas the highly cooperative (SSB)₃₅ state could be important for DNA replication (47). Although our present results pinpoint a role in DNA repair for the positive cooperative binding of plant Whirly proteins to ssDNA, both of these studies stresses the biological importance of the cooperative binding of ssDNA-binding proteins to DNA. Another similarity between plant Whirlies and SSB proteins is the dependence of both higher order protein assembly and cooperative DNA binding on positively charged residue. Indeed, for the SSB proteins from *E. coli* and *T. aquaticus*, the mutation of an arginine residue to alanine results in a lower protein assembly and to a decrease in the cooperative binding to DNA (49). Overall, these studies outline additional similarities between plant Whirlies and SSB proteins although the mechanistic details differ between these two classes of proteins.

AFM was used on several instances to monitor a change in the oligomeric state of proteins upon DNA binding. For example, AFM revealed that APOBEC3G, a ssDNA cytosine deaminase that mainly exists as a monomer in solution, forms dimers upon ssDNA binding (50). Likewise, CbpA, a nucleoid-associated DNA-binding protein which in free form exists in a monomer-dimer equilibrium, was shown to form large aggregates upon binding long double-stranded DNA molecules (51). The present study reveals that upon ssDNA binding, tetramers of WHY2 form superstructures that resemble the 24-mers observed in the F432 crystal space group. The comparison of WHY2 with other DNA

binding proteins monitored by AFM reveals that WHY2 does not form large aggregates upon DNA binding. However, WHY2 still induces DNA condensation as WHY2-DNA complexes appear as packed beads whereas SSB-DNA complexes appear as filaments with well-defined contour (38). This packing could be mediated by protein-protein interactions in between adjacent 24-mers.

The 24-mer assembly of plant Whirly proteins closely resemble that of ferritin (52), sulfur oxygenase reductase (53) and that of the heat shock proteins HSP16.5 and DegP (54,55). In each case, the assembly of 24 protomers generates a hollow protein shells with 432 symmetry. The outer diameter of the shell varies from 12 nm for HSP16.5 to 19.5 nm for DegP. For all these nanostructures, pores or chimneys that lead to the interior of the sphere generally coincide with symmetry axes and Whirly proteins are no exceptions. Indeed, WHY2 possesses six 0.8 nm-wide pores that coincide with the four-fold axes that gain access to a 5 nm diameter hollow core. A recent study on ferritin demonstrated that the mutation of single residues, termed oligomerization switch residues, is sufficient to abolish the formation of the spherical assembly (56). Our data indicates that K67 could fulfill a similar role for WHY2.

The peculiar assembly of plant Whirly proteins into 24-mers could promote binding to both DNA and metabolites. Indeed, most of the proteins that form nanostructures with 432 symmetry also store compounds. Ferritin is known to store iron atoms (52,57), the enzyme sulfur oxygenase reductase stores metabolites (53) whereas the heat shock proteins HSP16.5 and DegP both store proteins (54,55). One possibility would be that plant Whirly proteins bind and encapsulate harmful chemical compounds thereby protecting DNA against damage. In this regard, it is interesting that DPS (DNA-protecting Protein during Starvation), an homolog of ferritin that also forms a hollow protein shell, protects DNA from oxidative stress by binding DNA and by sequestering iron in its core (57). It is also worth mentioning that ferredoxin:sulfite reductase, an enzyme that localizes to chloroplast nucleoids, also binds concomitantly DNA and sulfites (58). It would therefore be possible that Whirly proteins protect DNA by sequestering harmful molecules.

In a recent report, Kwon et al. confirmed that DNA DSBs are repaired in *Arabidopsis* chloroplasts through a microhomology-mediated DNA repair pathway and brought mechanistic insights into the repair process (59). Of particular interest is the fact that a single DSB is repaired using microhomologies located more than 3 kb apart from each other. This suggests that during the repair process, very long stretches of ssDNA are unveiled before finding matching microhomologous DNA sequences. Through 24-mer assembly and positive cooperative binding, Whirly proteins could bind and protect these regions as well as the single-stranded regions which are abundantly present elsewhere in plant organelles genomes (60). This would enable faithful repair of DNA and prevent MMBIR product accumulation. Thus, our new findings complements and extends the model we previously proposed (15).

In conclusion, our data highlight a role of the second lysine of the KGKAAL motif of plant Whirly proteins in both the tolerance against DNA damages and higher order protein assembly. Further work is however required to determine precisely how the 24-mer assembly of plant Whirly proteins is involved in DNA repair. Also, it will be interesting to verify whether the 24-mer assembly of Whirly proteins is involved in metabolite binding and what is the functional consequence of this interaction.

4.7 References

1. Puchta, H. (2005) *J Exp Bot* **56**, 1-14
2. Kimura, S., and Sakaguchi, K. (2006) *Chem Rev* **106**, 753-766
3. Nielsen, B. L., Cupp, J. D., and Brammer, J. (2010) *J Exp Bot* **61**, 2535-2537
4. Marechal, A., and Brisson, N. (2010) *New Phytol* **186**, 299-317
5. Rhoads, D. M., Umbach, A. L., Subbaiah, C. C., and Siedow, J. N. (2006) *Plant Physiol* **141**, 357-366
6. Apel, K., and Hirt, H. (2004) *Annu Rev Plant Biol* **55**, 373-399
7. Palmer, J. D. (1983) *Nature* **301**, 92-93
8. Lonsdale, D. M., Brears, T., Hodge, T. P., Melville, S. E., and Rottmann, W. H. (1988) *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **319**, 149-163
9. Pang, Q., Hays, J. B., and Rajagopal, I. (1992) *Proc Natl Acad Sci USA* **89**, 8073-8077
10. Cerutti, H., Osman, M., Grandoni, P., and Jagendorf, A. T. (1992) *Proc Natl Acad Sci USA* **89**, 8068-8072
11. Khazi, F. R., Edmondson, A. C., and Nielsen, B. L. (2003) *Mol Genet Genomics* **269**, 454-463
12. Cerutti, H., and Jagendorf, A. T. (1993) *Plant Physiol* **102**, 145-153
13. Manchekar, M., Scisum-Gunn, K., Song, D., Khazi, F., McLean, S. L., and Nielsen, B. L. (2006) *J Mol Biol* **356**, 288-299
14. Rowan, B. A., Oldenburg, D. J., and Bendich, A. J. (2010) *J Exp Bot* **61**, 2575-2588
15. Cappadocia, L., Marechal, A., Parent, J. S., Lepage, E., Sygusch, J., and Brisson, N. (2010) *Plant Cell* **22**, 1849-1867
16. Khakhlova, O., and Bock, R. (2006) *Plant J* **46**, 85-94
17. Hastings, P. J., Ira, G., and Lupski, J. R. (2009) *PLoS Genet* **5**, e1000327
18. Desveaux, D., Maréchal, A., and Brisson, N. (2005) *Trends Plant Sci* **10**, 95-102
19. Desveaux, D., Allard, J., Brisson, N., and Sygusch, J. (2002) *Nat Struct Biol* **9**, 512-517

20. Yoo, H. H., Kwon, C., Lee, M. M., and Chung, I. K. (2007) *Plant J* **49**, 442-451
21. Prikrýl, J., Watkins, K. P., Friso, G., Wijk, K. J., and Barkan, A. (2008) *Nucleic Acids Res* **36**, 5152-5165
22. Maréchal, A., Parent, J. S., Véronneau-Lafortune, F., Joyeux, A., Lang, B. F., and Brisson, N. (2009) *Proc Natl Acad Sci USA* **106**, 14693-14698
23. Krause, K., Kilbiński, I., Mulisch, M., Rödiger, A., Schäfer, A., and Krupinska, K. (2005) *FEBS Lett* **579**, 3707-3712
24. Maréchal, A., Parent, J. S., Sabar, M., Véronneau-Lafortune, F., Abou-Rached, C., and Brisson, N. (2008) *BMC Plant Biol* **8**, 42
25. Edmondson, A. C., Song, D., Alvarez, L. A., Wall, M. K., Almond, D., McClellan, D. A., Maxwell, A., and Nielsen, B. L. (2005) *Mol Genet Genomics* **273**, 115-122
26. Horvath, M. P. (2008) Single-stranded Nucleic Acid-binding Proteins. in *Protein-Nucleic Acid Interactions: Structural Biology* (Rice, P. A., and Correll, C. C. eds.), Royal Society of Chemistry, London. pp 91-128
27. Kim, C., and Wold, M. S. (1995) *Biochemistry* **34**, 2058-2064
28. Lohman, T. M., and Ferrari, M. E. (1994) *Annu Rev Biochem* **63**, 527-570
29. Raghunathan, S., Kozlov, A. G., Lohman, T. M., and Waksman, G. (2000) *Nat Struct Biol* **7**, 648-652
30. Cappadocia, L., Sygusch, J., and Brisson, N. (2008) *Acta Crystallogr Sect F Struct Biol Cryst Commun* **64**, 1056-1059
31. Zheng, L., Baumann, U., and Reymond, J. L. (2004) *Nucleic Acids Res* **32**, e115
32. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501
33. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr D Biol Crystallogr* **54**, 905-921
34. Brunger, A. T. (2007) *Nat Protoc* **2**, 2728-2733
35. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S.,

- Terwilliger, T. C., and Zwart, P. H. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221
36. DeLano, W. L. (2002) *DeLano Scientific, San Carlos, CA, USA*
 37. Argaman, M., Golan, R., Thomson, N. H., and Hansma, H. G. (1997) *Nucleic Acids Res* **25**, 4379-4384
 38. Hamon, L., Pastré, D., Dupaigne, P., Le Breton, C., Le Cam, E., and Piétrement, O. (2007) *Nucleic Acids Res* **35**, e58
 39. Hellens, R. P., Edwards, E. A., Leyland, N. R., Bean, S., and Mullineaux, P. M. (2000) *Plant Mol Biol* **42**, 819-832
 40. Clough, S. J., and Bent, A. F. (1998) *Plant J* **16**, 735-743
 41. Krissinel, E., and Henrick, K. (2007) *J Mol Biol* **372**, 774-797
 42. Lohman, T. M., Overman, L. B., and Datta, S. (1986) *J Mol Biol* **187**, 603-615
 43. Reichel, C., Maas, C., Schulze, S., Schell, J., and Steinbiss, H. H. (1996) *J Gen Virol* **77 (Pt 4)**, 587-592
 44. Haseltine, C. A., and Kowalczykowski, S. C. (2002) *Mol Microbiol* **43**, 1505-1515
 45. Mumtsidu, E., Makhov, A. M., Konarev, P. V., Svergun, D. I., Griffith, J. D., and Tucker, P. A. (2008) *J Struct Biol* **161**, 172-187
 46. Meyer, R. R., and Laine, P. S. (1990) *Microbiol Rev* **54**, 342-380
 47. Carlini, L. E., Porter, R. D., Curth, U., and Urbanke, C. (1993) *Mol Microbiol* **10**, 1067-1075
 48. Ferrari, M. E., Fang, J., and Lohman, T. M. (1997) *Biophys Chem* **64**, 235-251
 49. Witte, G., Fedorov, R., and Curth, U. (2008) *Biophys J* **94**, 2269-2279
 50. Shlyakhtenko, L. S., Lushnikov, A. Y., Li, M., Lackey, L., Harris, R. S., and Lyubchenko, Y. L. (2010) *J Biol Chem*
 51. Cosgriff, S., Chintakayala, K., Chim, Y. T., Chen, X., Allen, S., Lovering, A. L., and Grainger, D. C. (2010) *Mol Microbiol* **77**, 1289-1300
 52. Theil, E. C. (1987) *Annu Rev Biochem* **56**, 289-315
 53. Urich, T., Gomes, C. M., Kletzin, A., and Frazão, C. (2006) *Science* **311**, 996-1000
 54. Krojer, T., Sawa, J., Schäfer, E., Saibil, H. R., Ehrmann, M., and Clausen, T. (2008) *Nature* **453**, 885-890
 55. Sutter, M., Boehringer, D., Gutmann, S., Günther, S., Prangishvili, D., Loessner, M. J., Stetter, K. O., Weber-Ban, E., and Ban, N. (2008) *Nat Struct Mol Biol*

56. Zhang, Y., Raudah, S., Teo, H., Teo, G. W., Fan, R., Sun, X., and Orner, B. P. (2010) *J Biol Chem* **285**, 12078-12086
57. Grant, R. A., Filman, D. J., Finkel, S. E., Kolter, R., and Hogle, J. M. (1998) *Nat Struct Biol* **5**, 294-303
58. Sekine, K., Fujiwara, M., Nakayama, M., Takao, T., Hase, T., and Sato, N. (2007) *FEBS J* **274**, 2054-2069
59. Kwon, T., Huq, E., and Herrin, D. L. (2010) *Proc Natl Acad Sci U S A* **107**, 13954-13959
60. Backert, S., Lurz, R., Oyarzabal, O. A., and Börner, T. (1997) *Plant Mol Biol* **33**, 1037-1050
61. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) *Nucleic Acids Res* **35**, W375-383

4.8 Footnotes

The atomic coordinates and structure factors for WHY2 K67A form I, WHY2 K67A form II and WHY2 K67A-dT₃₂ will be deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

The assistance of Jihyun Daniel Yi and Antonio Nanci for help with AFM is gratefully acknowledged. This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada and from the Fonds québécois de la recherche sur la nature et les technologies to both N.B. and J.S. and from the Canadian Institutes of Health Research to J.S. Work was carried out in part at beamline X29 of the National Synchrotron Light Source. Financial support comes principally from the Offices of Biological and Environmental Research and of Basic Energy Sciences of the US Department of Energy, and from the National Center for Research Resources of the National Institutes of Health. Some of the research was performed at the 08ID-1 beamline at the Canadian Light Source. Assistance by beamline personnel is gratefully acknowledged.

The abbreviations used are: AFM, atomic force microscopy; DSBs, DNA double-strand breaks; EMSA, Electrophoretic mobility shift assays; MMBIR, microhomology-mediated break induced replication; RMSD, root-mean square deviation; SSB, single-stranded DNA-binding protein.

Table 4-I. Data collection and refinement statistics

Protein	WHY2 K67A Free Form I	WHY2 K67A Free Form II	WHY2 K67A-dT₃₂
Beamline/Wavelength	CLS-08ID-1/0.98 Å	NSLS-X29/1.29 Å	NSLS-X29/1.08 Å
Space group	F432	P4 ₂ 2	F432
Cell dimensions : <i>a</i> , <i>b</i> , <i>c</i> (Å)	168.37, 168.37, 168.37	88.36, 88.36, 51.70	166.47, 166.47, 166.47
Resolution (Å)	50–2.35 (2.48–2.35)	50–1.78 (1.84–1.78)	50–2.45 (2.54–2.45)
Total Reflections	201296	258074	135360
Unique Reflections	9012	20443	7685
R _{sym} (%)	5.9 (102.5)	4.8 (31)	6.7 (67)
I / σ I	31.0 (2.8)	21.5 (2.6)	24.5 (2.3)
Completeness (%)	100.0 (100.0)	99.4 (95.3)	99.6 (97.8)
Multiplicity	22.3 (18.6)	12.6 (5.8)	17.6 (7.0)
Refinement Statistics			
Resolution (Å)	50–2.35	50–1.78	50–2.45
Reflections (total/test)	9012/541	19650/960	7469/586
R _{work} / R _{free} (%)	22.36/26.64	19.14/21.66	24.98/26.73
Number of atoms			
Protein	1263	1209	1263
DNA/Phosphate	10	10	180
Water	42	209	39
B-factors			
Protein	60.8	37.2	68.1
DNA/Phosphate	69.0	48.9	75.4
Water	60.1	49.4	66.2
RMSDs			
Bond lengths (Å)	0.005	0.007	0.003
Bond angles (°)	0.894	1.001	0.705
Ramachandran^a			
Favored (%)	98.1	98.7	94.3
Outliers (%)	0	0	0

Values in parentheses are for highest-resolution shell.

^a MolProbity analysis (59)

Figure 4-1. Critical role of K67 in the assembly of tetramers into 24-mers. *A*, Left, surface representation of the 24-mer assembly, obtained by applying crystallographic symmetry along the 4- 3- and 2-fold axes, in the free form structure of WHY2 (PDB 3N1H). Individual tetramers are colored in different colors. Right, interaction between K67 and the backbone of F138 in the same structure. Protein residues are in stick representation. Hydrogen bonds are represented as black dashes. Top left, schematic of the 24 subunits drawn as spheres. *B*, Left, surface representation of the 24-mer assembly in the WHY2-ERE₃₂ complex structure (PDB 3N1I). The nucleotides are in stick representation. Right, interaction between K67 and T3 in the same structure. The presentation and orientation is similar as in panel A. *C*, Top, overall view of two adjacent 24-mers in the crystal of WHY2-ERE₃₂. The hexamers, depicted in surface representation are colored in green and yellow. Bottom, interface between two 24-mers showing the close vicinity of DNA fragments. DNA molecules are in stick representation. The black dash lines illustrate the two possible routes for the DNA. Top left, schematic of the two 24-mers with individual subunits drawn as spheres.

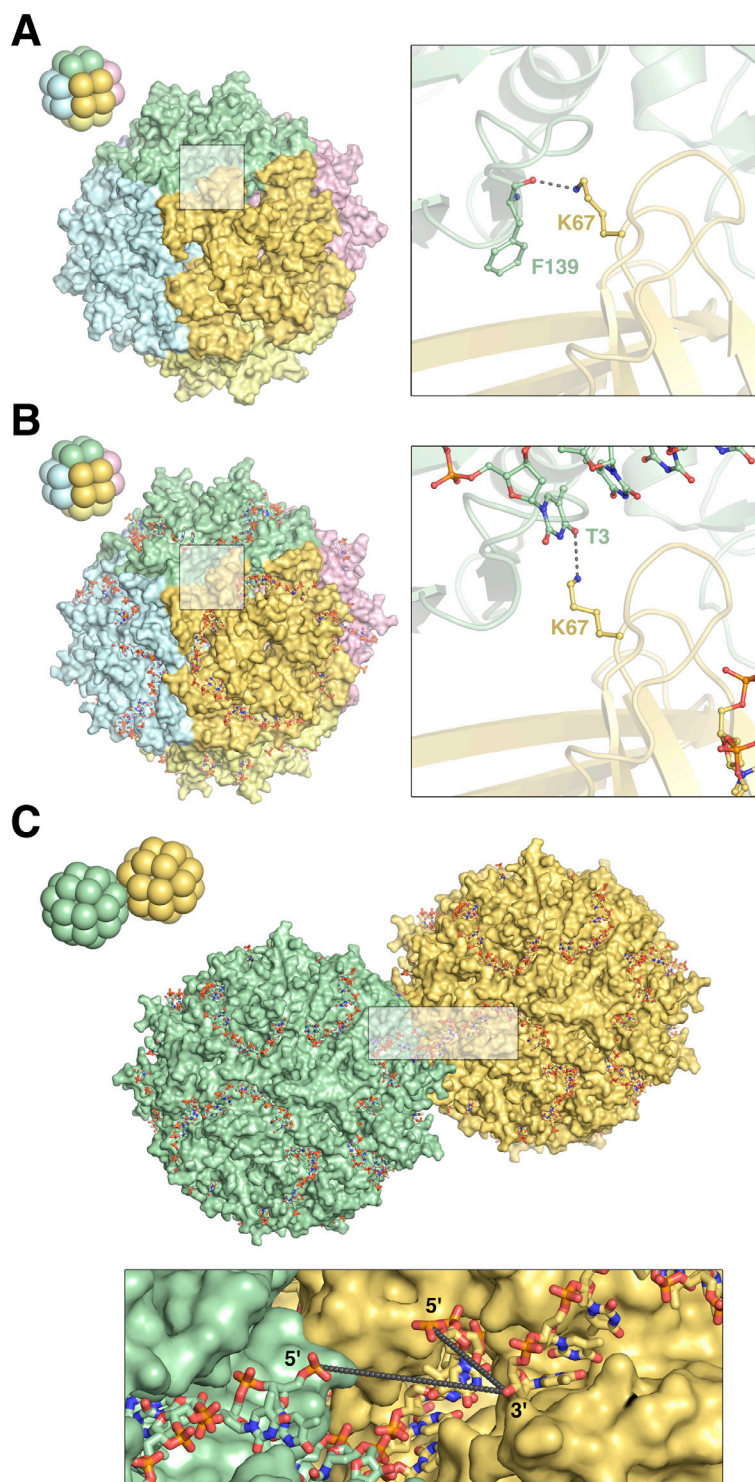


Figure 4-2. Whirly proteins form 24-mers in vitro upon binding long ssDNA molecules.

A, Topographic imaging of WHY2 in the free form on a mica surface obtained by AFM (left) and height measurement of the particles (right). *B*, Topographic imaging of WHY2-dT₃₂ complexes on a mica surface obtained by AFM (left) and height measurement of the particles (right). *C*, Topographic imaging of WHY2-M13mp18 complexes on a mica surface obtained by AFM (left) and height measurement of the particles (right). *D*, Phase imaging of WHY2-M13mp18 complexes on a mica surface obtained by AFM (left) and zoom on WHY2-M13mp18 complexes (right). All acquisitions were performed in tapping mode in air.

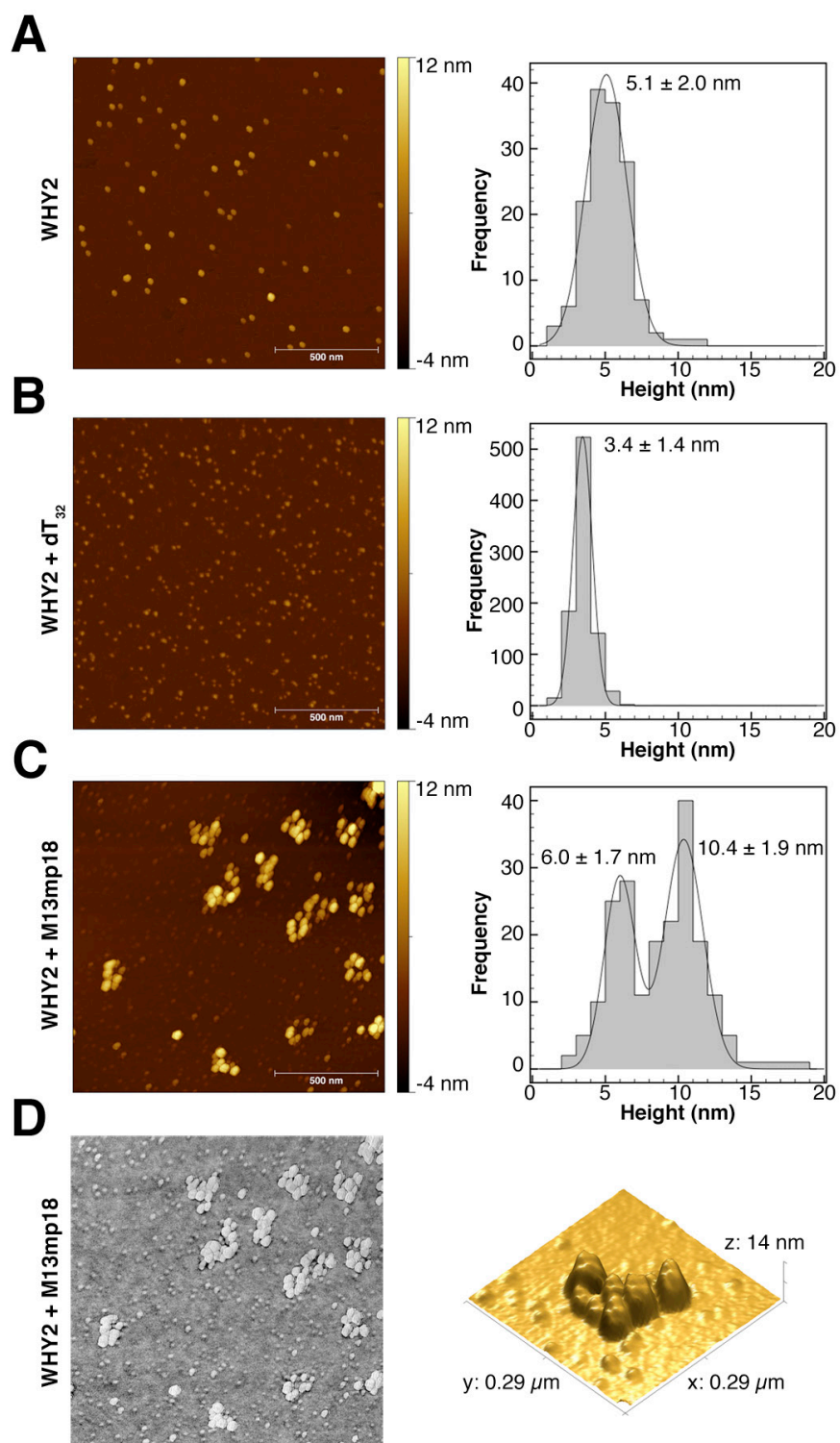
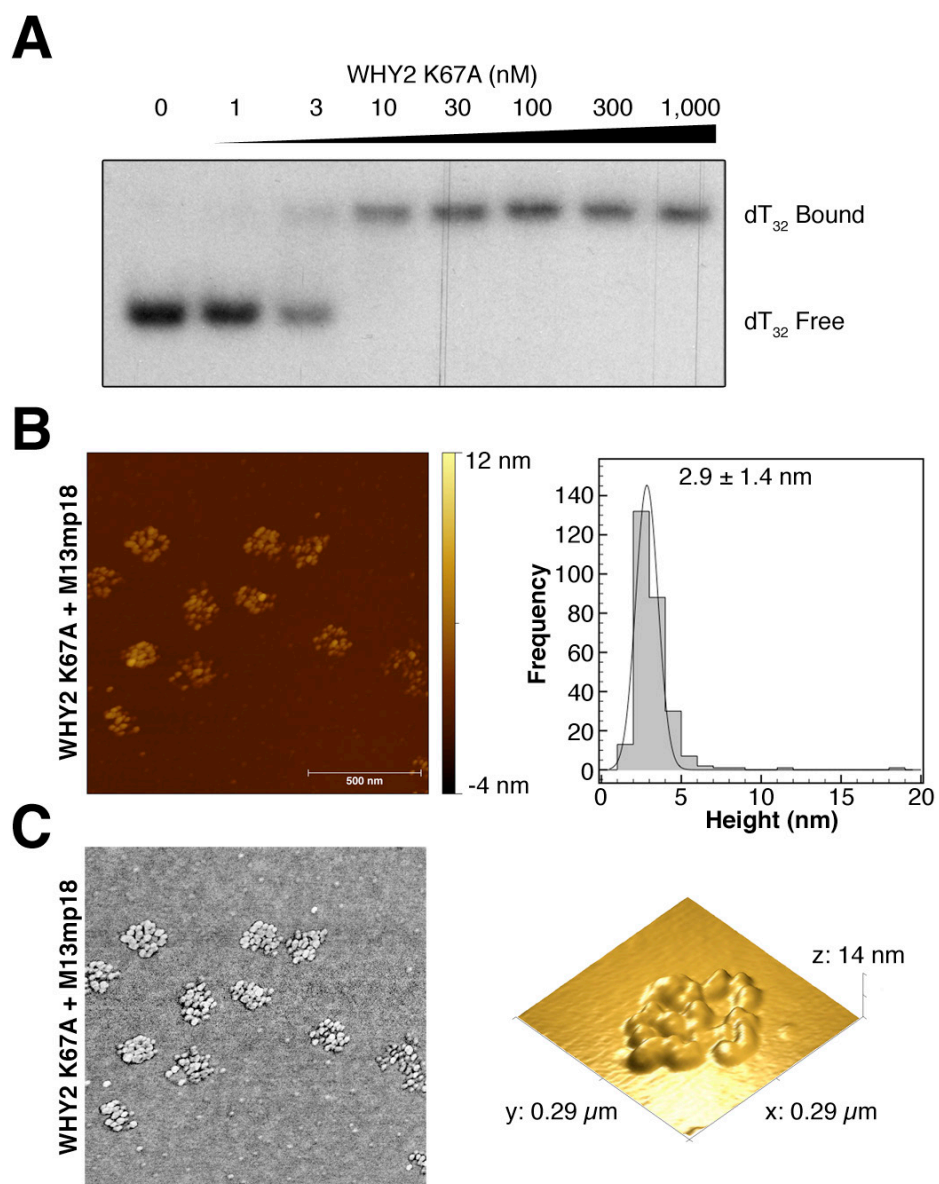


Figure 4-3. WHY2 K67A mutant does not assemble into 24-mers *in vitro*. *A*, Representative EMSA results showing the binding of WHY2 K67A to (dT)₃₂. Increasing amounts of WHY2 were incubated with target oligonucleotides and the complexes resolved on a 10% (w/v) polyacrylamide gel. *B*, Topographic imaging of WHY2 K67A-M13mp18 complexes on a mica surface obtained by AFM (left) and height measurement of the particles (right). *C*, Phase imaging of WHY2-M13mp18 complexes on a mica surface obtained by AFM (left) and zoom on WHY2-M13mp18 complexes (right). All acquisitions were performed in tapping mode in air.



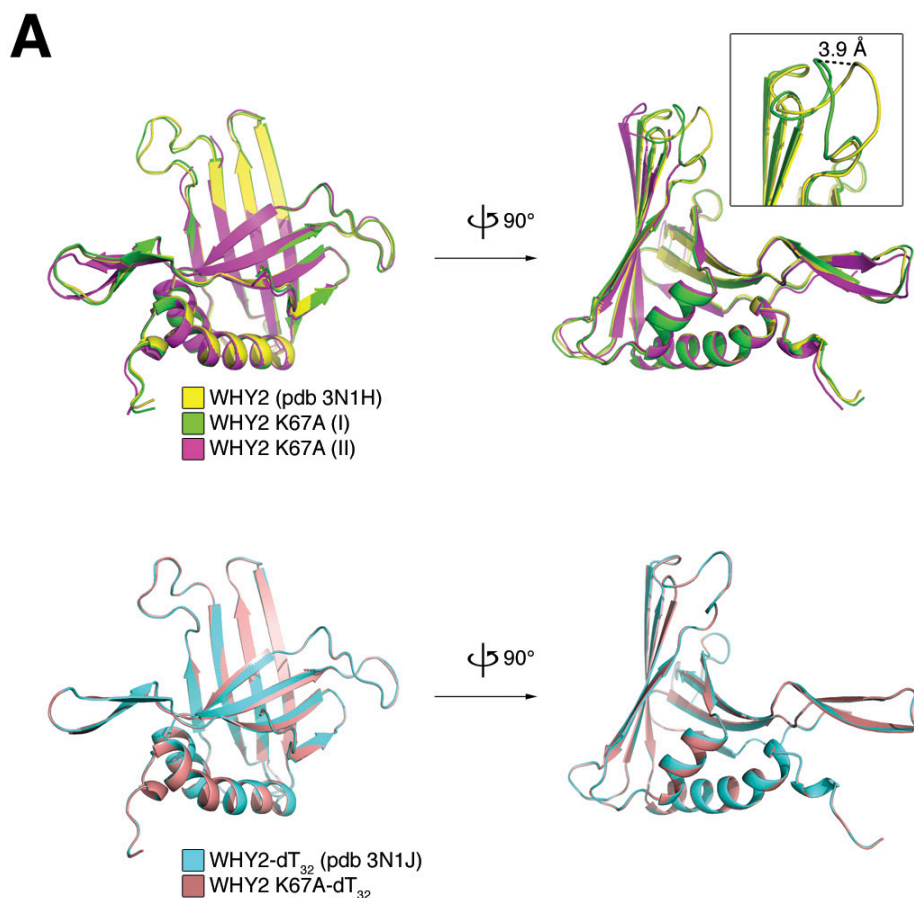


Figure 4-4. Structural comparison of the K67A variant with the nonmutated WHY2 protein. *A*, Comparison of the two crystal forms of WHY2 K67A with the nonmutated protein. The rmsd for superimposing the crystal forms I and II of WHY2 K67A onto the nonmutated protein are 0.7 Å and 1.1 Å, respectively. *B*, Comparison of the dT₃₂-bound form of WHY2 K67A with the dT₃₂-bound nonmutated protein. The rmsd for superimposing WHY2 K67A onto the nonmutated protein is 0.2 Å.

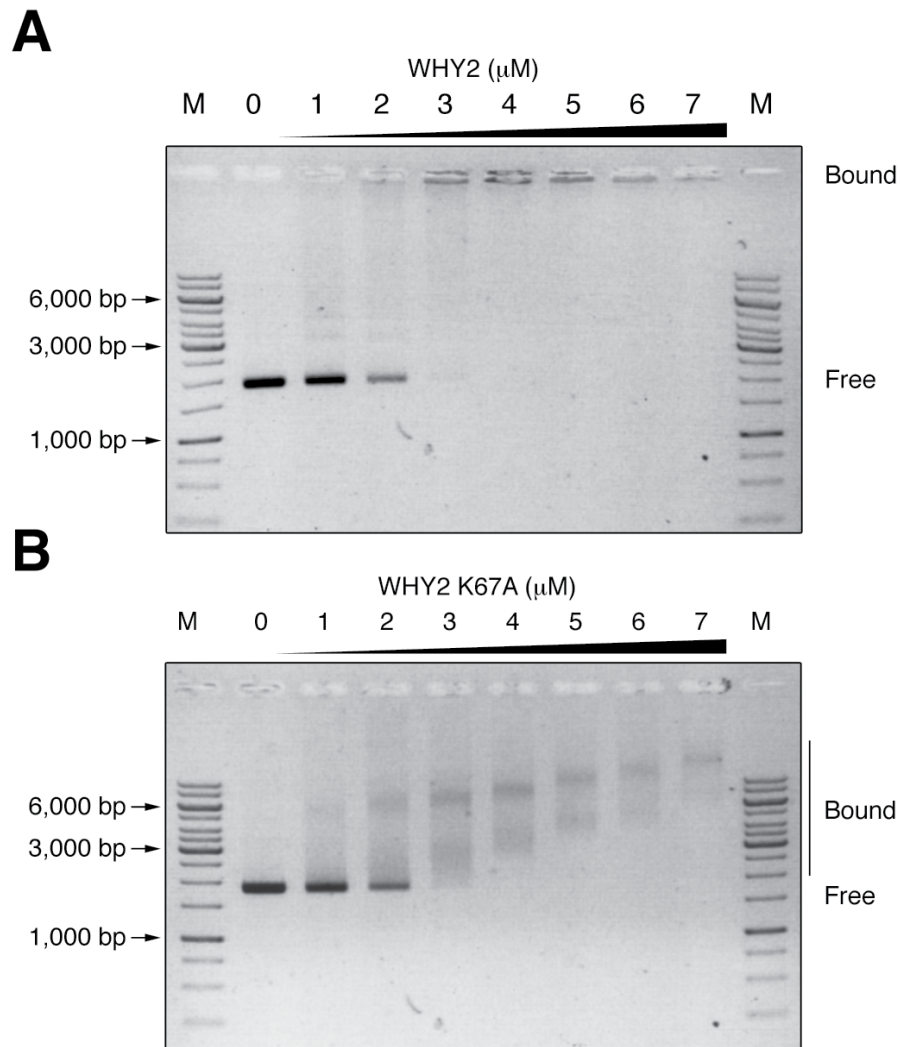


Figure 4-5. Cooperative binding of WHY2 upon binding long ssDNA molecules. *A*, Black/white inverted image of an agarose-based EMSA showing cooperative binding of WHY2 to M13mp18. *B*, Black/white inverted image of an agarose-based EMSA showing noncooperative binding of WHY2 K67A to M13mp18. The DNA was incubated with increasing amount of WHY2 or WHY2 K67A and the complexes were resolved on a 0.7% (w/v) agarose gel containing ethidium bromide. DNA and protein-DNA complexes were visualized by UV transillumination. M represents the molecular weight markers. “Bound” indicates the position of the protein-DNA complexes that barely entered the gel. “Free” indicates the unbound DNA.

Figure 4-6. WHY1 but not WHY1 K91A partially complements ciprofloxacin sensitivity in *why1 why3* background. *A*, Histogram displaying the percentage of plants with etiolated/variegated first true leaves after 10 days of growth on MS medium supplemented with 0.125 μ M ciprofloxacin. Error bars represent the standard deviation of the counts made on different plates. For this experiment, the entire sequence of WHY1 or WHY1 K91A under the control of the cauliflower mosaic virus 35S promoter was introduced into *why1why3* plants. Except for the WT control, two independent homozygous lines were used for each genotype. A complete statistical treatment of these data can be found in Table 4-SI. *B*, The protein level of WHY1/3 was assessed in each line by protein gel blot. Whirly proteins were detected by using an α -WHY1/3 antibody. A section of the blot stained with Ponceau red is presented below as a loading control. *C*, The ssDNA-binding activity of Whirly proteins was monitored by electrophoretic mobility shift assay using crude plastid protein extracts isolated from plants of the indicated genotypes and a radiolabeled dT₃₂ oligonucleotide. A section of an SDS-PAGE stained with Coomassie blue is presented below as a loading control.

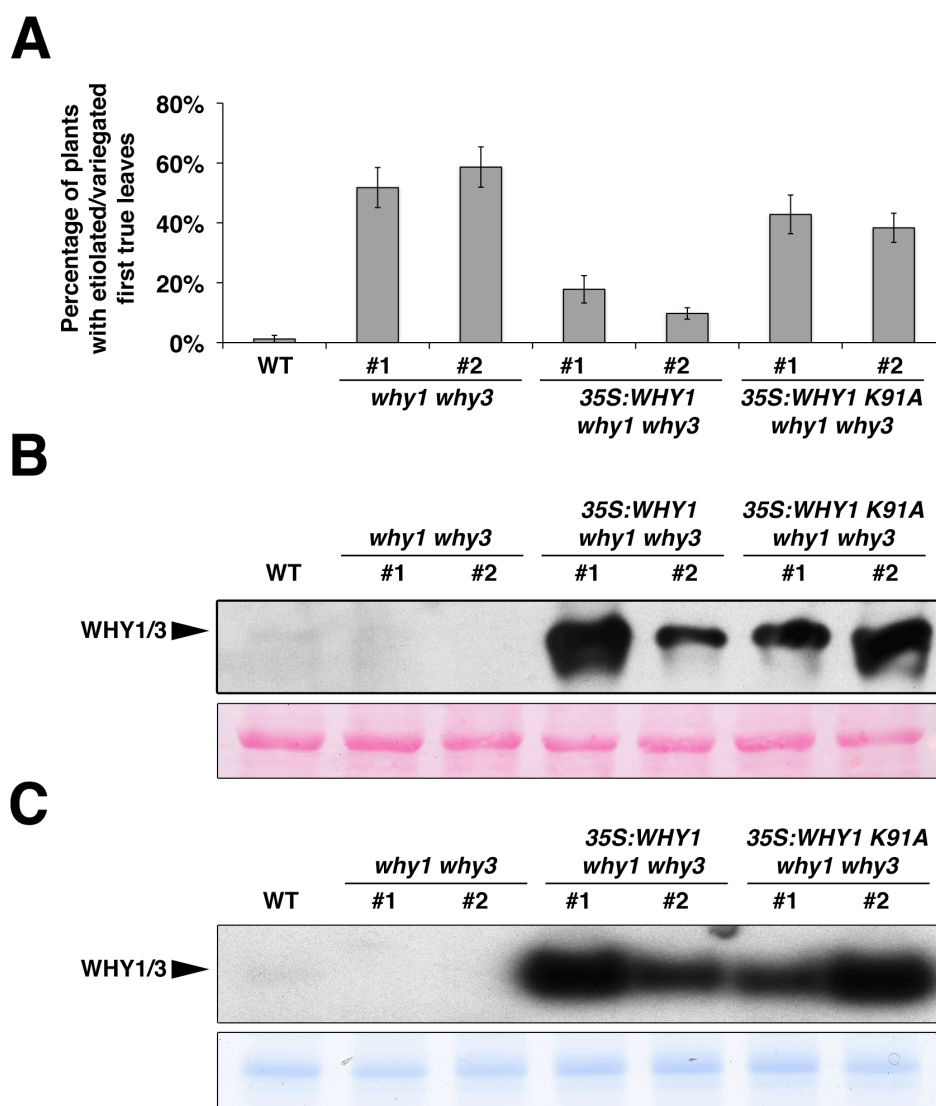


Figure 4-S1. Surface representation of the hexameric assembly of Whirly tetramers in the free form structure of WHY2 (PDB 3N1H). The hollow sphere is viewed along the four-fold axis. The molecular surface is color-coded by electrostatic potential with potentials of less than -40kT colored in red, neutral potentials (0kT) colored in white and potentials greater than +40kT colored in blue. The electrostatic potential of the protein was computed with APBS and visualized using PyMol (DeLano Scientific).

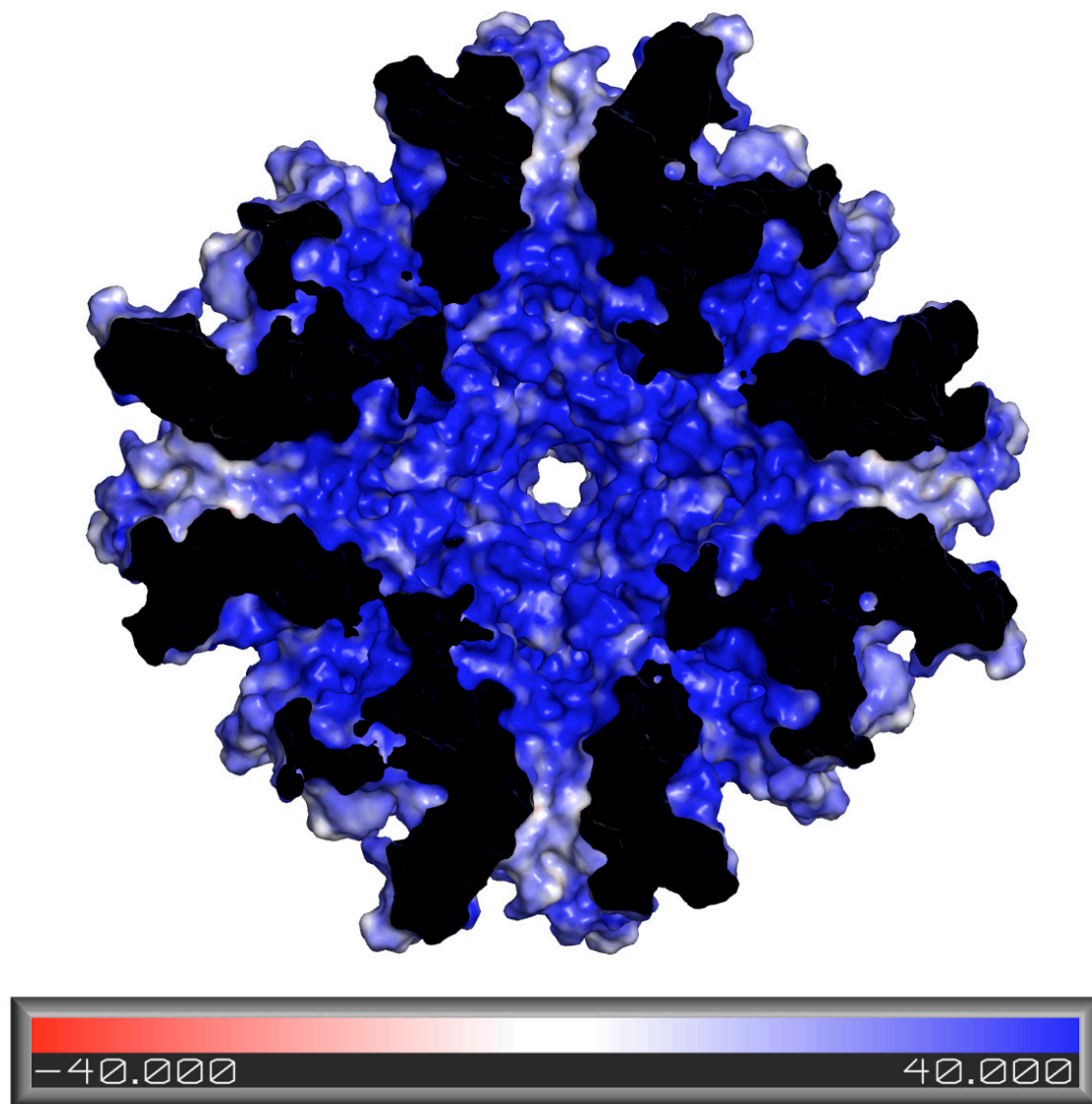


Figure 4-S2. *A.* Size exclusion chromatography of WHY2 and WHY2 K67A in the free form. An apparent molecular weight of 61 kDa is consistent with the formation of tetramers (predicted molecular weight of 80 kDa). *B.* Size exclusion chromatography of WHY2 and WHY2 K67A in complex with a dT₃₂ oligonucleotide. An apparent molecular weight of 84 kDa is consistent with the formation of DNA-bound tetramers (predicted molecular weight of 90 kDa). For each sample, 500 µl of protein or complex at a 5 mg/ml concentration was injected on a Superdex 200 10/300 column equilibrated with 50 mM Tris/HCl pH8 and 250 mM NaCl at a 0.75 ml/min flow rate. The absorbance of the solution was measured at 280 nm. The apparent molecular weights were computed following calibration of the column using the Molecular Weight Marker Kit (Sigma-Aldrich). 1, bovine thyroglobulin (669 kDa); 2, horse spleen apoferritin (443 kDa); 3, sweet potato β-amylase (200 kDa); 4, yeast alcohol dehydrogenase (150 kDa); 5, bovine serum albumin (66kDa); 6, bovine erythrocytes carbonic anhydrase (29 kDa); 7, horse heart cytochrome C (12 kDa).

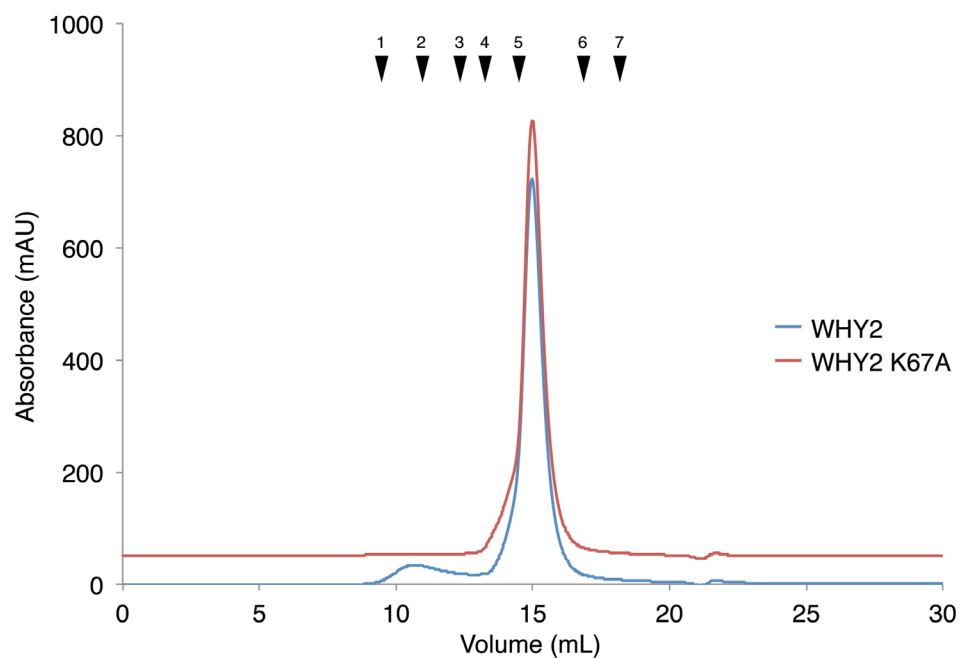
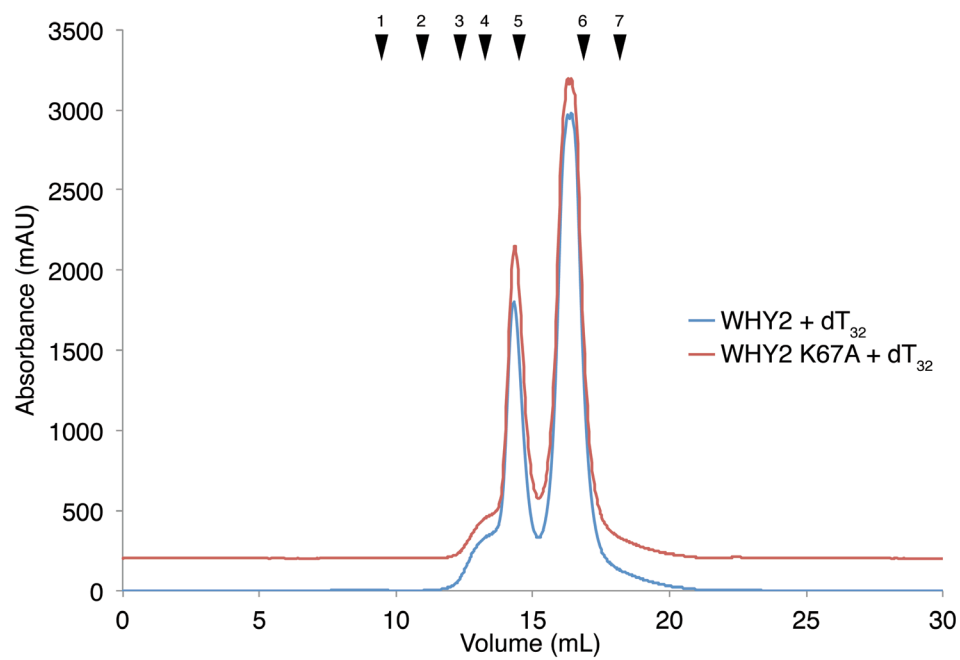
A**B**

Table 4-SI. Statistical analysis of the complementation of the etiolation/variegation phenotype. One-way ANOVA and Tukey's Honestly Significant Difference statistical tests were used for multiple phenotype comparisons. —, * and ** denote statistical significance at $P > 0.05$, $P < 0.05$ and $P < 0.001$.

Genotype	Col-0	<i>why1 why3</i> #1	<i>why1 why3</i> #2	<i>35S:WHY1</i> <i>why1 why3</i> #1	<i>35S:WHY1</i> <i>why1 why3</i> #2	<i>35S:WHY1</i> K91A <i>why1 why3</i> #1	<i>35S:WHY1</i> K91A <i>why1 why3</i> #2
Col-0	—	**	**	**	—	**	**
<i>why1 why3</i> #1	**	—	—	**	**	—	*
<i>why1 why3</i> #2	**	—	—	**	**	**	**
<i>35S:WHY1</i> <i>why1 why3</i> #1	**	**	**	—	—	**	**
<i>35S:WHY1</i> <i>why1 why3</i> #2	—	**	**	—	—	**	**
<i>35S:WHY1</i> K91A <i>why1 why3</i> #1	**	—	**	**	**	—	—
<i>35S:WHY1</i> K91A <i>why1 why3</i> #2	**	*	**	**	**	—	—

CHAPITRE 5

Discussion

Le but de ce projet de recherche est de comprendre comment les protéines Whirly de plantes interagissent avec l'ADN monocaténaire et d'en apprendre davantage sur les conséquences biologiques de cette interaction. Dans cette section, nous allons discuter les implications de nos recherches pour la radiocristallographie des macromolécules, la biologie structurale des protéines Whirly et la biologie des organites de plantes.

5.1 Implications pour la radiocristallographie

La première partie de notre travail a consisté en l'obtention de cristaux d'un complexe Whirly-ADN diffractant à haute résolution. Faire croître des cristaux de la molécule d'intérêt est, rappelons-le, une étape difficile, aléatoire, mais *indispensable* dans le processus de résolution de structure tridimensionnelle par radiocristallographie. Le processus se fait par essai-erreur et les cas problématiques font fréquemment l'objet de publications de manière à aider la communauté des cristallographes en proposant des pistes de dépannage.

Nous avons ainsi commencé nos travaux avec des cristaux d'un complexe WHY1-ADN qui ne diffractaient qu'à faible résolution. En utilisant différents oligonucléotides, en optimisant plusieurs paramètres (température de croissance des cristaux, pH, concentration en sel et en agent précipitant de la solution de cristallisation, nature et concentration de l'agent de cryoprotection, etc.) et en modifiant la construction de WHY1 de manière à supprimer les régions flexibles de la protéine, il nous a été possible d'améliorer le potentiel de diffraction des cristaux du complexe. Néanmoins, la diffraction de ces cristaux était extrêmement anisotrope et ne permettait pas l'élucidation de la structure du complexe. Un criblage à haut débit n'a pas non plus permis d'identifier de nouvelles conditions de cristallisation. Ce n'est qu'en utilisant une stratégie de criblage d'homologues que nous avons pu obtenir la structure d'un complexe Whirly-ADN et débiter notre étude structurale.

Le criblage de protéines homologues (orthologues ou paralogues) est une stratégie permettant d'augmenter les chances de cristallisation d'un type de protéine donné (Pour revue, se référer à (189)). Cette stratégie repose sur le fait *(i)* que les acides aminés importants pour la fonction de la protéine sont habituellement conservés entre des espèces voisines. *(ii)* que les acides aminés situés en surface de la protéine sont généralement moins bien conservés que

le cœur hydrophobe de la protéine qui est nécessaire au bon repliement de la protéine (190), (iii) que ce sont les acides aminés situés en surface de la protéine qui permettent l'assemblage de la protéine au sein d'une maille cristalline et (iv) que la mutation de certains résidus de surface peut favoriser l'assemblage de la protéine dans une maille cristalline (191). Dans le cas du domaine Whirly qui compte 160 résidus, 77 résidus varient entre WHY1 et WHY2. Notre étude montre que ces changements n'influencent pas la structure globale de la protéine mais permettent la cristallisation et l'élucidation de la structure dans le cas de WHY2.

5.2 Implications pour la biologie structurale des protéines Whirly

5.2.1 Un nouveau mode de liaison à l'ADN monocaténaire

Comme pour les autres protéines liant l'ADN monocaténaire, WHY2 utilise des interactions hydrophobes, des ponts hydrogènes, des empilements aromatique ainsi que des interactions électrostatiques afin de stabiliser l'ADN. Une particularité unique au mode de liaison des protéines Whirly de plantes est cependant que l'ADN monocaténaire est pris en sandwich entre des sous-unités contiguës. Ainsi, contrairement à de nombreuses protéines liant l'ADN monocaténaire, les protéines Whirly de plantes n'utilisent pas des domaines liant individuellement l'ADN, mais des domaines qui, collectivement, lient l'ADN. La liaison ne peut donc se faire que si la protéine adopte une conformation minimalement tétramérique. À cet égard, on remarque que cet agencement est stabilisé par très nombreuses interactions hydrophobes. Celles-ci préviennent la dissociation du tétramère et sont si stables qu'au cours de nos recherches il ne nous a jamais été possible de détecter la forme monomérique de la protéine en condition non dénaturante. Ce modèle de liaison à l'ADN est aussi différent de celui proposé lors de l'élucidation de la structure de WHY1 (140). Celui-ci prévoyait en effet que la liaison s'effectuerait au cœur des planchers β de la protéine et que l'espace séparant les planchers β adjacents servirait soit de point de passage soit de point d'entrée/sortie de l'ADN.

5.2.2 Liaison à l'ADN et à l'ARN

Deux études récentes suggèrent qu'en plus de l'ADN monocaténaire, les protéines Whirly lient également l'ARN (171,172). Pour vérifier si notre modèle est conciliable avec ces

résultats, nous y avons remplacé l'ADN monocaténaire par de l'ARN. Une telle modélisation n'occasionne que peu d'encombrement stérique, suggérant ainsi que l'ARN pourrait être lié dans une conformation très voisine de celle de l'ADN. D'un point de vue quantitatif, nous avons effectué des expériences de retard sur gel en utilisant des sondes rERE₃₂ et rU₃₂ qui ont respectivement la même longueur et la même séquence que les sondes ERE₃₂ et dT₃₂ décrites au chapitre 3. Ces sondes sont liées avec environ trente fois moins d'affinité qu'un ADN monocaténaire de même séquence. Ceci est en accord avec les travaux du Dr. Barkan qui démontrent que les Whirly lient tout l'ADN chloroplastique mais seulement quelques ARN (171).

5.2.3 Spécificité de liaison à l'ADN monocaténaire

Plusieurs études ont caractérisé l'interaction des protéines Whirly de plantes aux acides nucléiques. Celles-ci sont résumées dans le tableau 5-I qui indique, pour chaque étude, les techniques et les protéines employées. Deux tendances sont à remarquer. Premièrement, la grande majorité des études ont montré que les protéines Whirly de plantes peuvent se lier, au moins *in vitro*, à l'ADN monocaténaire. En plus de l'ADN monocaténaire et de l'ARN, les protéines Whirly pourraient aussi lier de l'ADN bicaténaire (162,168,171) quoiqu'avec une affinité substantiellement plus faible (162,171). Deuxièmement, les études qui se sont intéressées à la liaison des protéines Whirly dans les organites ont montré une liaison soit à des ADN dégénérés (192), soit à de très nombreuses séquences d'ADN dans le génome des organites (171,173,193). En revanche, celles qui se sont intéressées à l'interaction des protéines Whirly avec des séquences nucléaires ont constaté une liaison séquence spécifique à l'élément ERE (162), à l'élément KPRE (168) ou à de l'ADN télomérique (170). Ces séquences, bien que riches en A/T, ne présentent pas de similarités les unes par rapport aux autres, si bien qu'aucune séquence consensus de liaison n'a pu à ce jour être déterminée.

Nos travaux montrent que les protéines Whirly peuvent lier des ADN de différentes séquences avec une affinité semblable et selon un même mode de liaison. La comparaison des structures de St-WHY2 en complexe avec différents ADN permet en outre de mettre en évidence comment la protéine s'adapte à la séquence qui lui est présentée. Nos résultats sont ainsi cohérents avec une liaison de manière séquence-indépendante dans les organites.

Tableau 5-I. Spécificité de liaison des protéines Whirly de plantes aux acides nucléiques. Pour chaque protéine Whirly étudiée, la spécificité est indiquée. La méthode ayant servi à l'analyse est également indiquée.

Protéine Whirly étudiée	Spécificité de liaison	Méthode(s) employée(s)	Référence(s)
At-WHY1	Multiples séquences d'ADN chloroplastique	ChIP	Maréchal, 2009
At-WHY1	ADN simple-brin télomérique : (TTTAGGG) _n	CA, RG	Kwon, 2004 Yoo, 2007
At-WHY1	ADN simple et double brin : élément KPPE	CA, RG, ChIP	Xiong, 2009
At-WHY2	Multiples séquences d'ADN mitochondrial	ChIP	Maréchal, 2009
At-WHY3	ADN simple et double brin : élément KPPE	CA, RG, ChIP	Xiong, 2009
Hv-WHY1	ADN simple brin : élément ERE et séquence télomérique	?	Grabowski, 2008 (données non publiées)
Hv-WHY1	ADN nucléaire et chloroplastique ainsi que quelques ARN chloroplastiques	RT-ChIP; IL	Melonek, 2010
St-WHY1	ADN simple brin : élément ERE (TGACAnnnnTGTCa)	CA, RG, ChIP	Desveaux, 2000 Desveaux, 2004 González-Lamothe, 2008
St-WHY2	ADN simple-brin de séquence dégénérée	CA	Vermel, 2002
Zm-WHY1	Multiples séquences d'ADN chloroplastique ainsi que quelques ARN chloroplastiques	RT-ChIP, SGS	Prikryl, 2008

ChIP, Immunoprécipitation de la chromatine; CA, Chromatographie d'affinité; RG, Retard sur gel; SGS, Sédimentation sur gradient de sucrose; RT-ChIP, Immunoprécipitation suivie de la transcription réverse et de l'hybridation sur une puce; IL, immunolocalisation

Pour expliquer la liaison plus spécifique des Whirly à certaines séquences nucléaires, nous proposons plusieurs hypothèses : (i) la forte teneur en A/T des éléments nucléaires pourrait leur conférer une plus grande propension à adopter une conformation monocaténaire ce qui faciliterait leur liaison par les protéines Whirly. (ii) Au niveau du noyau, les protéines Whirly pourraient interagir avec d'autres protéines de liaison à l'ADN et ainsi être recrutés au niveau de certains éléments d'ADN. (iii) Étant donné que les Whirly semblent moins abondantes au niveau du noyau qu'au niveau des organites (174), il est possible qu'elles ne puissent se lier qu'à un nombre restreint de sites avec lesquels elles ont plus d'affinité. Notons que des préférences de séquences ont été observées chez les protéines liant l'ADN monocaténaire de manière séquence-indépendante. Ainsi, la protéine SSB d'*Escherichia coli* possède une nette préférence pour des oligomères constitués de thymines plutôt que d'adénines (156). Aussi, de tous les ADN que nous avons utilisés, c'est l'oligonucléotide dont la séquence est basée sur l'élément ERE qui est lié avec le plus d'affinité par St-WHY2. Ceci suggère que les protéines Whirly, lorsque présents en quantité non saturante, pourraient se lier de manière préférentielle à certains sites.

5.2.4 Compétition entre l'ADN et le motif acide/aromatique C-terminal

La structure cristallographique de WHY1 présente une interaction entre les feuillets β de la protéine et le motif acide/aromatique C-terminal qui se replie en hélice α (140). Lorsque l'on compare cette structure avec celle d'un complexe WHY2-ADN, on remarque que le motif acide/aromatique C-terminal interagit avec des acides aminés qui sont également impliqués dans la liaison à l'ADN monocaténaire (Figure 5-1). Ceci suggère l'existence d'une compétition entre ce motif et l'ADN pour la liaison aux feuillets β de la protéine. Cette idée est corroborée par le fait qu'un mutant de délétion de WHY1 ne possédant pas de motif acide/aromatique à son extrémité C-terminale possède une plus grande affinité pour l'ADN monocaténaire (181).

5.2.5 Agencement en 24-mères

Dans les cristaux de WHY2, en forme libre où en complexe avec de l'ADN monocaténaire, les protéines sont non seulement agencées en tétramères, mais également

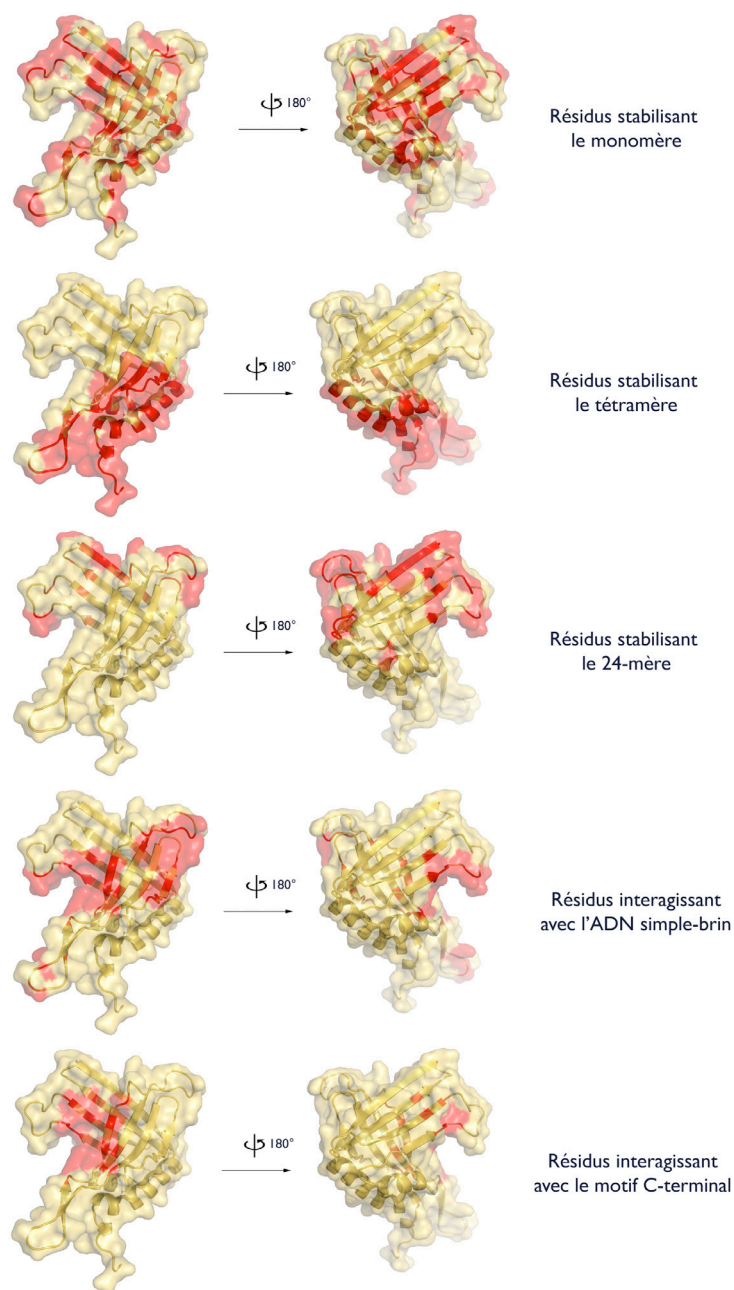


Figure 5-1. Représentation d'un monomère de WHY2 en style surface et cartoon. Les résidus importants pour l'interaction avec l'ADN monocaténaire, l'interaction avec le motif acide/aromatique C-terminal, la stabilisation du tétramère ou la stabilisation du 24-mère sont colorés en rouge, respectivement.

en 24-mères formant ainsi des particules de 480 kDa. Le serveur PISA (194), qui détecte les agencements stables de protéines au sein des mailles cristallines, prédit que cette conformation pourrait exister, au moins *in vitro*. Les 24-mères ont une symétrie de type 432 et forment des sphères de douze nanomètres de diamètre possédant une cavité intérieure de cinq nanomètres de diamètre avec un volume de 90 nm³. Des pores de 0,8 nm de diamètre, localisés au niveau des axes de symétrie d'ordre quatre, permettent au solvant d'accéder à cette cavité. La formation de 24-mères entraîne le recouvrement d'une surface de 3,3 nm², soit la moitié de la surface totale de la protéine. De cette surface, 2,2 nm², soit 65% de l'espace total recouvert, sont dédiés à la formation de tétramères, tandis que les 1,1 nm² restant sont dédiés à l'agencement des tétramères en 24-mères. Chaque monomère effectue onze ponts d'hydrogène de manière à former des tétramères, et sept de plus afin de former des 24-mères.

Globalement, l'agencement en 24-mère de WHY2 est similaire à celui des protéines de choc thermique HSP16.5 et DegP, de la réductase/oxygénase du soufre, de la ferritine ou de la déhydratase de l'imidazoleglycérol phosphate (Figure 5-2). Nos travaux montrent en outre que cet agencement est métastable *in vitro*, étant donné qu'il est induit lors de la cristallisation ou lors de la liaison de WHY2 à de longues séquences d'ADN. La lysine 67, qui est conservée parmi les protéines Whirly de plantes, joue un rôle très important dans cet agencement, étant donné que la mutation de ce résidu en alanine empêche la formation de 24-mères lors de la liaison de longues séquences d'ADN. Une étude sur la bacterioferritine a montré que la mutation de quelques résidus commutateurs, ou switch residues, est suffisante pour empêcher la formation de 24-mères (195). Nos travaux attribuent un rôle semblable à la lysine 67 de WHY2 tout en confirmant que ce résidu n'est ni important pour la stabilité du tétramère ni requis pour la liaison à de courtes séquences d'ADN.

L'agencement en 24-mères permet de mieux comprendre le mode de liaison à l'ADN des protéines Whirly. En effet, nous avons remarqué dans nos modèles cristallographiques que WHY2 stabilise l'ADN aux extrémités des feuillettes β et non en leur centre. En fait, à cet endroit, l'ADN peut soit transiter d'une protéine à l'autre au sein d'un tétramère ou transiter d'un 24-mère à l'autre. Nous voyons donc que le centre des feuillettes β n'est pas seulement un point de passage, mais plutôt un point d'aiguillage permettant de diriger l'ADN et de construire un réseau nucléoprotéique.

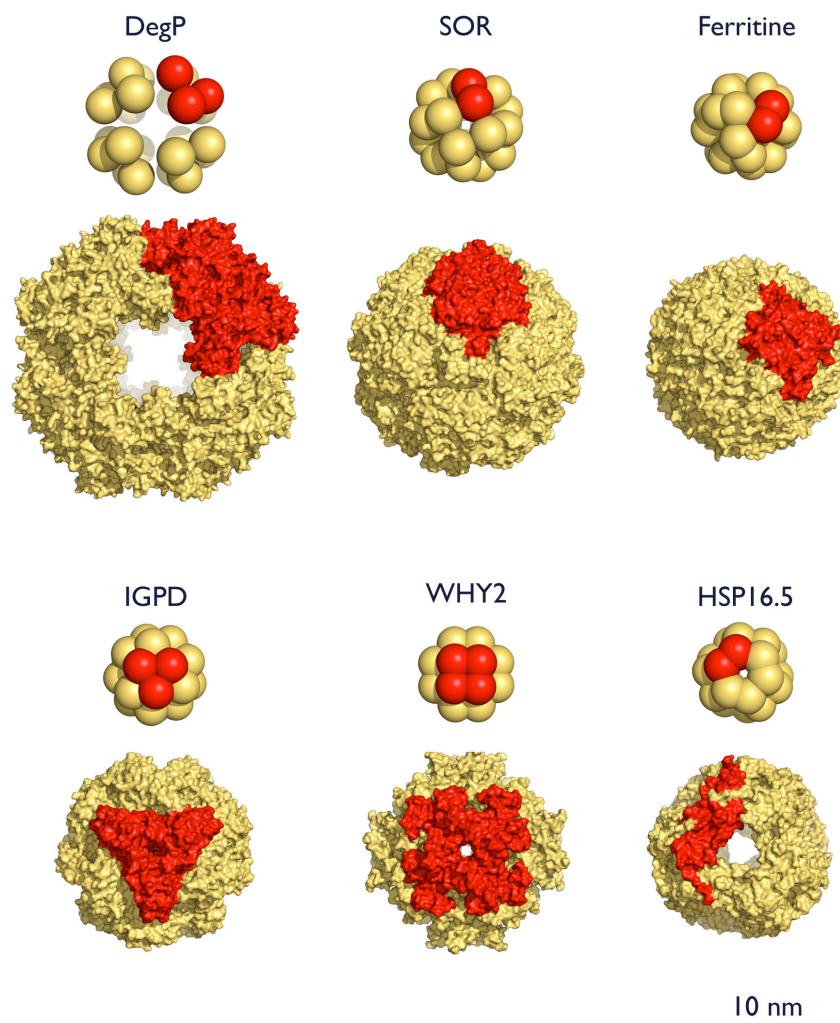


Figure 5-2. Représentation en style surface de protéines s'assemblant en 24-mères. Au dessus de chaque protéine se trouve un schéma montrant l'organisation des monomères au sein du 24-mère. Chaque 24-mère résulte de l'assemblage de protomères qui sont colorés en rouge. DegP, protéine de choc thermique DegP de la bactérie *Escherichia coli* (code pdb 3CS0). SOR, Oxygénase/réductase du soufre de l'archée *Acidianus ambivalens* (code pdb 2CB2). Ferritine, ferritine de la mitochondrie humaine (code pdb 1R03). IGPD, déshydratase d'imidazoleglycérol phosphate d'*Arabidopsis* (code pdb 2F1D). WHY2, Whirly 2 de la plante *Solanum tuberosum* (code pdb 3N1H). HSP16.5, protéine de choc thermique HSP16.5 de l'archée *Methanococcus jannaschii* (code pdb 1SHS).

5.2.6 Liaison coopérative à l'ADN

Notre étude révèle que la liaison des protéines Whirly à l'ADN monocaténaire s'effectue de manière coopérative. Un premier type de coopérativité peut-être mis en évidence dans les expériences de liaison entre les protéines Whirly et de courts ADN monocaténaires. En effet, en analysant nos expériences de retards sur gel, nous avons calculé un coefficient de Hill de $2,9 \pm 0,3$ pour la liaison de WHY2 à dT₃₂ et un coefficient de $3,3 \pm 0,5$ pour la liaison de WHY2 K67A à dT₃₂. Cette coopérativité positive semble liée au fait que trois sous-unités du tétramère sont nécessaires pour pouvoir lier dT₃₂ (en comptant dix nucléotides par sous-unité). Conformément à nos résultats montrant que la liaison de courtes séquences d'ADN n'est pas altérée chez le mutant WHY2 K67A, la mutation de K67A n'entraîne pas de changement significatif dans le coefficient de Hill. Un deuxième type de coopérativité peut être mis en évidence dans les expériences de liaison entre protéines Whirly et les longs ADN monocaténaires. En effet, par une approche de retard sur gel d'agarose, nous avons mis en évidence une différence qualitative entre la liaison de WHY2 à l'ADN M13mp18 et la liaison de WHY2 K67A à l'ADN M13mp18.

De nombreuses protéines liant l'ADN monocaténaire avec peu de spécificité de séquence le font avec une coopérativité positive (pour revue, se référer à (120,196)). Dans tous les cas connus, cette coopérativité est permise grâce à la formation d'un filament nucléoprotéique lors de la liaison de longues séquences d'ADN. Dans le cas des protéines Whirly, c'est également la formation d'une superstructure qui permet la liaison coopérative à l'ADN. Cependant, il ne s'agit pas d'un filament nucléoprotéique mais plutôt d'un réseau de 24-mères. Ce réseau confère aux complexes WHY2-M13 une apparence compacte.

Le rôle biologique de la coopérativité positive des protéines liant l'ADN monocaténaire est, encore aujourd'hui, mal connu. D'un point de vue théorique, la coopérativité positive permet le recouvrement plus rapide et mieux ordonné de l'ADN, empêchant ainsi la formation de structures secondaires, les événements de recombinaison illégitimes, ainsi que la dégradation de l'ADN par des nucléases. Une étude portant sur la protéine SSB d'*Escherichia coli* suggère que le mode (SSB)₆₅, qui lie l'ADN avec une faible coopérativité positive, est impliqué dans la réparation de l'ADN tandis que le mode (SSB)₃₅, qui lie l'ADN avec une forte coopérativité positive, est plutôt impliqué dans la réplication de l'ADN. Nos

travaux suggèrent également que la liaison coopérative à l'ADN aurait un rôle biologique dans le cas des protéines Whirly. Cependant, nous observons plutôt une corrélation entre la liaison coopérative à l'ADN et la réparation de l'ADN.

5.2.7 Résidus importants

La conservation stricte de certains résidus du domaine Whirly peut s'expliquer par leur rôle dans au moins une des fonctions suivantes : (i) maintien de la structure du monomère, (ii) formation de tétramères, (iii) formation de 24-mères, (iv) liaison aux acides nucléiques, (v) interaction avec le motif acide/aromatique C-terminal, (vi) interaction avec d'autres protéines, et (vii) régulation par des modifications post-traductionnelles. Nos données structurales permettent d'identifier les résidus impliqués dans les cinq premières fonctions (Figure 5-1). De plus, on remarque que certains résidus peuvent jouer plusieurs rôles. Par exemple, l'histidine 174 peut lier un des phosphates de l'ADN tout en stabilisant un des planchers β de la protéine de par sa liaison avec le glutamate 122. La lysine 188 peut également lier un phosphate de l'ADN tout en stabilisant le même plancher β de par sa liaison avec le groupement carbonyle de la leucine 115. Nos données structurales permettent ainsi de comprendre pourquoi certains résidus sont si bien conservés alors que d'autres le sont moins.

5.2.8 Ressemblances et différences structurales

Nos résultats montrent une similarité de fonction entre les protéines SSB bactériennes et les protéines Whirly. Bien que ces protéines aient une organisation semblable et que des feuilletts β soient impliqués dans la liaison à l'ADN dans les deux cas, elles ne possèdent pas de similarité de séquences ou de structure (Score Z de 1,1, signifiant qu'il n'existe pas de similarité de structure significative entre les deux protéines analysées d'après le serveur DaliLite (197)).

Les protéines MRP1/2 possèdent quant à elles une topologie ainsi qu'un agencement tridimensionnel très proche de ceux des protéines Whirly de plantes (141). De plus, plusieurs des acides aminés impliqués dans la formation du cœur hydrophobe sont semblables pour ces deux classes de protéines. Ceci suggère que ces protéines ont évolué à partir d'un ancêtre

commun. Cependant, elles possèdent des mécanismes différents de liaison aux acides nucléiques, et les acides aminés impliqués dans la liaison aux acides nucléiques ne sont pas localisés aux mêmes endroits. Ces analyses sont cohérentes avec le fait que ces protéines aient des rôles cellulaires différents et elles démontrent que le domaine Whirly est suffisamment versatile pour pouvoir être adapté dans deux contextes biologiques différents.

Récemment, la structure de la protéine DdrB de la bactérie *Deinococcus radiodurans* a été élucidée (198). Cette protéine forme un pentamère avec une symétrie de type C₅ (chaque sous-unité est relié à ses voisines par une rotation centrale de 72°). Chaque monomère possède un domaine semblable au OB-fold qui est responsable de la liaison à l'ADN monocaténaire. Les auteurs proposent que l'ADN s'enroulerait autour de la protéine (198). Dans cette conformation, le complexe aurait un aspect similaire à celui formé entre St-WHY2 et l'ADN monocaténaire. Il faudra cependant attendre la résolution de la structure d'un complexe protéine/ADN pour déterminer si la stratégie de liaison à l'ADN est similaire pour ces protéines et les Whirly.

5.3 Implications pour la biologie des organites de plantes

L'observation que des réarrangements de l'ADN chloroplastique apparaissent chez des plantes d'*Arabidopsis* n'exprimant pas les protéines WHY1 et WHY3 (173) nous a poussé à étudier la réparation de l'ADN dans les organites des plantes. Comme nous l'avons vu en introduction, peu de choses sont connues concernant les mécanismes de réparation de l'ADN dans les organites des cellules eucaryotes. Ceci est en partie dû au fait qu'il est difficile d'endommager spécifiquement l'ADN des organites de manière à en observer la réparation. Le premier système permettant de générer des cassures bicaténares dans les mitochondries n'a en effet été mis en place chez la souris qu'en 2001 (199). En 2008, on détecte pour la première fois la réparation de cassures bicaténares dans le génome mitochondrial de la drosophile suite au traitement de cellules en culture avec de la bléomycine (200). Cette même année, on détecte la réparation de cassures bicaténares dans le génome chloroplastique de l'algue *Chlamydomonas* suite à l'expression d'une endonucléase dirigée vers le chloroplaste (99).

Il n'existait pas encore de système permettant d'induire des cassures bicaténares de l'ADN dans les organites d'*Arabidopsis* au moment où nous avons débuté notre étude. Le premier pas à faire pour étudier la réparation de l'ADN dans les organites d'*Arabidopsis* consistait donc à élaborer un système permettant d'induire la formation de cassures bicaténares dans les organites. Deux étudiants de notre laboratoire, Étienne Lepage et Alexandre Maréchal, ont montré qu'il est possible d'utiliser certains inhibiteurs d'ADN gyrase, notamment la ciprofloxacine, pour induire des cassures bicaténares de l'ADN spécifiquement dans l'ADN des mitochondries et des chloroplastes de plantes. C'est ce système que nous avons utilisé pour étudier les mécanismes de réparation de l'ADN des organites de plantes.

5.3.1 Élaboration de mécanismes de réparation des cassures bicaténares de l'ADN dans les organites d'*Arabidopsis*

En traitant des plantes d'*Arabidopsis* avec la ciprofloxacine, il nous a été possible de démontrer que dans les chloroplastes et les mitochondries d'*Arabidopsis*, les cassures bicaténares de l'ADN sont réparées par un mécanisme très similaire au MMBIR observé chez l'humain et la levure. Cette réparation s'accompagne de l'apparition de réarrangements de l'ADN médiés des très courtes séquences répétées et disséminées dans le génome (les séquences répétées que nous avons isolé ont une taille comprise entre 5 et 37 nucléotides). Nous avons également montré qu'une recombinaison dépendante de courtes séquences répétées d'ADN permettait de réparer les cassures bicaténares de l'ADN dans les mitochondries d'*Arabidopsis*.

Une étude du Dr. Bendich effectuée parallèlement à nos recherches est venue corroborer notre approche expérimentale en montrant que des cassures bicaténares de l'ADN induites par la ciprofloxacine entraînent des changements dans la structure de l'ADN chloroplastique (201). Ceux-ci sont en outre exacerbés en l'absence de la recombinaison RecA1 qui possède un signal de localisation pour les chloroplastes. Aussi, peu de temps après la publication de notre étude, le groupe du Dr. Herrin a établi un système conditionnel permettant d'exprimer l'endonucléase I-CreII dans les chloroplastes d'*Arabidopsis* et ainsi d'induire une cassure bicaténale d'ADN à un endroit précis et connu du génome (106). Les plantes exprimant cette endonucléase présentent des phénotypes variables et certaines d'entre

elles sont variées. Des expériences de PCR montrent que les cassures sont réparées par un mécanisme de recombinaison dépendant de très courtes séquences répétées et disséminées, ce qui corrobore nos résultats. Il est intéressant de constater que ces séquences répétées sont localisées relativement loin (en général à plus de 1 000 nucléotides) du site de cassure. Ceci suggère que, lors de la réparation, de très longues séquences d'ADN monocaténaire sont dévoilées avant de trouver des séquences répétées compatibles. Les auteurs n'ont cependant pas vérifié si les cassures de l'ADN pouvaient mener à la formation de réarrangements autre que des délétions. Ainsi, ils proposent que l'ADN du chloroplaste serait réparé par un mécanisme de jonction d'extrémités dépendant de très courtes séquences d'ADN ou *microhomology-mediated end joining* (106).

5.3.2 Rôle des protéines Whirly dans la réparation des cassures bicaténares de l'ADN dans les organites d'*Arabidopsis*

Nos travaux ont permis de proposer deux modèles permettant d'expliquer le rôle des protéines Whirly dans la réparation des cassures bicaténares de l'ADN. Peu de temps après la cassure bicaténaires de l'ADN, il y a résection des extrémités 5' par des nucléases et exposition de potentiellement longues séquences d'ADN monocaténaire (106). Celles-ci pourraient alors être liées par les protéines Whirly de manière séquence-indépendante, empêchant ainsi l'appariement de cet ADN avec d'autres séquences dans le génome. La liaison des protéines Whirly à l'ADN endommagé pourrait aussi faciliter sa réparation grâce au recrutement de protéines impliquées dans la recombinaison homologue. Le motif acide/aromatique C-terminal des protéines Whirly pourrait permettre ce recrutement comme dans le cas des protéines SSB (202). Il est intéressant de noter que dans ce modèle une faible quantité de protéines Whirly est suffisante pour assurer la réparation de l'ADN, étant donné qu'une fois celle-ci achevée, les protéines Whirly sont relâchées.

Dans notre deuxième modèle, les protéines Whirly seraient constitutivement liées aux séquences d'ADN monocaténaire, qui représentent de cinq à dix pour cent du génome des organites de plantes (résultats non publiés de Étienne Lepage et Alexandre Maréchal et (203)). Ce faisant, les protéines Whirly pourraient prévenir l'attaque de ces régions par l'ADN monocaténaire provenant des régions endommagées, empêchant ainsi la formation de

réarrangements illégitimes et favorisant la réparation de l'ADN par recombinaison homologue. Ce modèle est cohérent avec le fait qu'une liaison constitutive des protéines Whirly à l'ADN des organites a été observée en l'absence de stress génotoxique (171,173,193). Ce modèle nécessite qu'une grande quantité de protéines Whirly soit présente dans les organites, de manière à protéger toutes les séquences d'ADN monocaténaire présentes dans le génome.

Globalement, nos deux modèles sont cohérents avec les propriétés biochimiques et structurales des protéines Whirly comme la liaison de manière séquence-indépendante à l'ADN ou la déstabilisation des duplex d'ADN. Nos expériences de retard sur gel montrent en effet qu'en plus de pouvoir lier des séquences différentes avec des affinités semblables, les protéines Whirly peuvent également dénaturer des ADN bicaténaire, limitant ainsi les risques d'appariement entre des séquences complémentaires. Ceci suggère que les protéines Whirly ont une action anti-recombinase. Ceci peut s'expliquer sur le plan structural si on compare le mode de liaison des protéines Whirly à celui de protéines dont la fonction est au contraire d'apparier des brins d'acides nucléiques comme RecA et de MRP1/2. Comme nous l'avons vu en introduction, ces protéines vont exposer les bases azotées des acides nucléiques monocaténaires de manière à en favoriser l'appariement. De plus, elles vont replier l'ADN ou l'ARN dans une conformation proche de celle adoptée par un ADN bicaténaire. Ce repliement s'effectue à l'échelle de toute la molécule d'ARN liée dans le cas de MRP1/2 tandis qu'il se fait un triplet d'ADN à la fois pour la protéine RecA. Contrairement à ces deux protéines, les Whirly vont prendre les bases azotées en sandwich, les empêchant ainsi de s'apparier avec des ADN de séquence complémentaire. De plus, la conformation dans laquelle se trouve l'ADN une fois lié par les protéines Whirly ne permet pas la formation d'un duplex d'ADN. La formation de 24-mère pourrait introduire un niveau de protection supplémentaire. En effet, lors de l'assemblage des protéines Whirly en sphères, il y a recouvrement partiel des bases azotées localisées entre les feuillets β de sous-unités contiguës. Ainsi, les protéines Whirly de plantes semblent avoir évoluées de manière à éviter la formation de duplex d'ADN. En ce sens, elles jouent un rôle opposé à celui des protéines MRP1/2 qui possèdent néanmoins une structure très proche.

5.4 Perspectives et travaux futurs

5.4.1 Corroborer des résultats

Plusieurs aspects de notre étude mériteraient une investigation plus poussée ainsi que des répétitions supplémentaires. En effet, bien que nos travaux suggèrent une conservation du mode de liaison à l'ADN ainsi que l'agencement en 24-mères entre les protéines Whirly mitochondriales et chloroplastiques, il serait bon dans le futur d'élucider la structure d'une protéine Whirly chloroplastique en complexe avec de l'ADN monocaténaire. Nous suggérons à cet égard d'utiliser uniquement le domaine Whirly et de tester la cristallisation de protéines provenant de différentes espèces (stratégie de criblage de protéines orthologues). Il serait également intéressant à l'avenir de concevoir de nouvelles mutations afin de déstabiliser l'agencement en 24-mères. Une fois les mutants obtenus, on pourrait vérifier si ceux-ci lient de longues séquences d'ADN monocaténaire sous une forme tétramérique et concevoir des mutations équivalentes dans des plantes d'*Arabidopsis*. Idéalement, on utiliserait le promoteur endogène des protéines Whirly plutôt que le promoteur 35S du virus de la mosaïque du chou-fleur. Finalement, nos travaux montrent une différence de comportement entre les lignées *why2-1* et *why2-2*. Il serait bon de vérifier si cette différence est réelle en retraitant les deux lignées à la ciprofloxacine et en remesurant le niveau de réarrangements de l'ADN.

5.4.2 Rôle des protéines Whirly dans la réplication de l'ADN

Nos travaux ont mis à jour une implication des protéines Whirly dans la réparation de l'ADN. Or, beaucoup de protéines liant l'ADN monocaténaire avec peu de spécificité de séquence comme Hs-RPA, Ec-SSB et T4 gp32 sont également impliquées dans la réplication de l'ADN (204). Aussi, dans les mitochondries des mammifères, la réplication de l'ADN semble dépendant de trois types de protéines : les polymérases γ , les hélicases TWINKLE et les protéines mtSSB (205). Les mtSSB permettent d'augmenter la processivité de l'hélicase TWINKLE ou de la polymérase γ , responsables de la réplication de l'ADN mitochondrial (206). Chez la drosophile, une réduction de leur niveau s'accompagne d'un retard de croissance ainsi que d'un appauvrissement en ADN mitochondrial (207). Ainsi, dans l'avenir, il serait bien de vérifier si les protéines Whirly (*i*) sont capables d'augmenter la

processivité des protéines impliquées dans la réplication de l'ADN et (ii) si elles interagissent physiquement ou génétiquement avec les protéines impliquées dans la réplication de l'ADN des organites de plantes.

5.4.3 Redondance fonctionnelle entre les protéines Whirly et les autres protéines liant l'ADN monocaténaire.

Nos résultats montrent que des plantes d'*Arabidopsis* n'exprimant pas la protéine WHY2 sont à peine plus sensibles aux cassures bicaténares de l'ADN que les plantes de type sauvage. Ceci suggère soit que WHY2 n'est qu'indirectement impliquée dans la réparation de l'ADN, soit que d'autres protéines pourraient relayer WHY2 dans les mitochondries. Les protéines mtSSB1/2 sont tout indiquées pour jouer ce rôle. Outre le fait que nous avons mis à jour une similarité fonctionnelle entre les protéines SSB bactériennes et les protéines Whirly, nous notons qu'il n'y a pas de protéine homologue aux protéines SSB dans les chloroplastes, où l'on observe un phénotype en absence des protéines Whirly. Il serait donc intéressant de vérifier par des croisements la relation épistatique entre WHY2 et mtSSB1/2. Outre les protéines mtSSB1/2, les protéines OSB, qui se localisent dans les chloroplastes et les mitochondries, pourraient être amenés à relayer la fonction des protéines Whirly. Ces études permettraient de clarifier les rapports existants entre ces protéines qui lient l'ADN avec peu de spécificité de séquence et de déterminer si certaines d'entre elles ont pu acquérir une spécialisation au cours de l'évolution, comme c'est le cas chez l'humain. Chez ce dernier, la protéine RPA est en effet impliquée dans la réplication, la recombinaison et la réparation de l'ADN, tandis que les protéines SSB1/2 se sont spécialisées dans la réparation de l'ADN (208).

5.4.4 Modulation de la conformation de l'ADN par les protéines Whirly

La liaison des protéines Whirly à l'ADN pourrait entraîner in vivo un changement dans la conformation de l'ADN. Nous avons en effet constaté que la formation de 24-mère par WHY2 entraîne la formation de complexes nucléoprotéiques compacts. Des protéines comme HU (Heat Unstable) lient l'ADN de manière séquence-indépendante et sont importantes pour la formation des nucléoïdes (pour revue, se référer à (209)). Leur liaison à l'ADN permet

en outre de moduler plusieurs fonctions cellulaires tel que la réplication, la recombinaison et la réparation de l'ADN (209). En plus de l'ADN bicaténaire, les protéines HU peuvent également lier l'ADN monocaténaire (210). Une analogie de fonction a même été proposée entre les protéines Whirly et les protéines HU (171). Pour tester cette hypothèse, on pourrait observer l'arrangement des génomes des organites de plantes en présence et en l'absence de protéines Whirly par électrophorèse en champ pulsé ou par microscopie électronique. On pourrait aussi s'intéresser à la conformation des nucléoïdes par microscopie confocale.

5.4.5 Rôle de l'agencement en 24-mères des protéines Whirly dans la réparation de l'ADN

Nos données suggèrent que l'agencement en 24-mères des protéines Whirly est important pour la fonction des protéines Whirly dans la réparation de l'ADN. En effet, l'introduction du gène *WHY1* dans des plantes d'*Arabidopsis* n'exprimant ni *WHY1* ni *WHY3* permet de compléter partiellement le phénotype de variégation, tandis que l'introduction du gène *WHY1* K91A ne le permet pas. Nous avons montré que cette mutation n'entraîne pas de perte de structure ou de stabilité de la protéine, ni d'altération dans la liaison de courtes séquences monocaténaires. Nos travaux ne nous permettent présentement pas de déterminer si l'absence de complémentarité du mutant *WHY1* K91A est due à la perte de la formation de 24-mère ou à la perte de la coopérativité de liaison à l'ADN. Des études de mutagenèse plus poussées devraient permettre de répondre à cette question.

5.4.6 Liaison de métabolites par les protéines Whirly

Il est intéressant de remarquer que les protéines ayant une symétrie de type 432 forment obligatoirement des sphères (211) et que la majorité de ces protéines sont impliquées dans la liaison de métabolites. En effet, les protéines de choc thermique DegP et HSP16.5 possèdent chacune une cavité qui encapsule les protéines permettant ainsi leur repliement ou leur dégradation (212,213). L'oxygénase/réductase du soufre et la déshydratase de l'imidazoleglycérol phosphate sont des enzymes qui ont respectivement pour substrat des

dérivés du soufre et l'imidazoleglycérol phosphate (214,215). La ferritine, quant à elle, permet la capture du fer, prévenant ainsi les dommages oxydatifs causés par la réaction de Fenton (216). Étant donné que les Whirly s'assemblent en 24-mères avec une symétrie 432, il serait possible qu'ils puissent eux aussi lier un métabolite. Un calcul de potentiel électrostatique révèle que l'intérieur de la cavité formée par WHY2 est électropositive, ce qui permettrait la liaison d'un composé chargé négativement. Ce composé pourrait diffuser à l'intérieur de la cavité centrale de WHY2 grâce aux pores localisés au niveau des axes de symétrie d'ordre 4.

Les protéines Whirly pourraient aussi se lier à des composés dangereux pour l'intégrité du génome. Ce faisant, elles joueraient un rôle similaire à la protéine bactérienne DPS (DNA-binding protein from starved cells). Cette protéine, qui est induite en condition de stress, possède la même architecture que la ferritine mais est agencée en dodécamère plutôt qu'en 24-mère (217). Elle lie et protège l'ADN bicaténaire, probablement en séquestrant le fer et en prévenant les dommages oxydatifs (217).

Une bonne communication entre le noyau et les organites est nécessaire afin de coordonner l'activité des cellules. Notamment, en cas de cassures bicaténaires, il est probable qu'un « message » soit envoyé des organites vers le noyau afin de stimuler la transcription des gènes impliqués dans la réparation de l'ADN. Bien qu'encore objet de nombreux débats, la communication rétrograde entre les organites et le noyau semble être dépendante de la variation de métabolites (218). Il est également possible que les Whirly, en se liant à des métabolites, altéreraient l'équilibre de ceux-ci. Ce débalancement, perçu par le noyau, permettrait la synthèse de protéines impliquées dans le métabolisme, et notamment dans la réparation de l'ADN.

5.5 Conclusion

Peu d'études se sont intéressées aux mécanismes de réparation du génome des organites de plantes. Toutefois, la compréhension de ces mécanismes est cruciale étant donné qu'une mauvaise réparation ou l'absence de réparation de dommages à l'ADN peut mener à une altération de l'information génétique. Si dans quelques cas ces changements peuvent être

bénéfiques pour l'humain (production de plantes ornementales variégues ou production de plantes présentant de la stérilité cytoplasmique mâle, utilisées pour la formation d'hybrides), dans la majorité des cas, ils sont soit sans conséquence visibles, soit nuisibles. Dans cette étude, nous avons montré que les protéines de la famille Whirly sont impliquées dans la réparation des cassures bicaténares de l'ADN dans les organites de plantes. Grâce à une approche de biologie structurale, nous avons pu expliquer comment la structure particulière des protéines Whirly, en forme de virevent, leur permet d'accomplir cette fonction. Notre approche structurale nous a également permis d'observer un nouvel état d'agencement quaternaire des protéines Whirly. Nos travaux font ainsi figure de pionniers dans l'étude des mécanismes de réparation de l'ADN des organites des plantes et suggèrent qu'il serait bon de caractériser d'autres protéines impliquées dans ce processus afin de mieux comprendre les mécanismes moléculaires de protection de l'information génétique extranucléaire.

Bibliographie

1. Evenari, M. (1989) *The botanical review* **55**, 106-139
2. Ernster, L., and Schatz, G. (1981) *J Cell Biol* **91**, 227s-255s
3. Wells, W. A. (2005) *J Cell Biol* **168**, 853
4. Sugiura, M. (2003) *Photosynth Res* **76**, 371-377
5. Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J., Staden, R., and Young, I. G. (1981) *Nature* **290**, 457-465
6. Shinozaki, K., Deno, H., Wakasugi, T., and Sugiura, M. (1986) *Curr Genet* **10**, 421-423
7. Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H., and Ozeki, H. (1986) **322**, 572-574
8. Puchta, H. (2005) *J Exp Bot* **56**, 1-14
9. Chatterjee, A., Mambo, E., and Sidransky, D. (2006) *Oncogene* **25**, 4663-4674
10. Lee, H. C., Chang, C. M., and Chi, C. W. (2010) *Ageing Res Rev* **9 Suppl 1**, S47-58
11. Pelletier, G., and Budar, F. (2007) *Curr Opin Biotechnol* **18**, 121-125
12. Sakamoto, W., Uno, Y., Zhang, Q., Miura, E., Kato, Y., and Sodmergen. (2009) *Plant Cell Physiol* **50**, 2069-2083
13. Dyllal, S. D., Brown, M. T., and Johnson, P. J. (2004) *Science* **304**, 253-257
14. Martin, W. (2003) *Proc Natl Acad Sci U S A* **100**, 8612-8614
15. Ravi, V., Khurana, J. P., Takagi, A. K., and Khurana, P. (2008) *Plant Systematics and Evolution* **271**, 101-122
16. Woodson, J. D., and Chory, J. (2008) *Nat Rev Genet* **9**, 383-395
17. Mackenzie, S. A. (2005) The mitochondrial genome of higher plants: A target for natural adaptation. in *Diversity and Evolution of Plants - Genotypic and phenotypic variation in higher plants* (Henry, R. ed.), CABI Publishers, London. pp 69-80
18. Day, A., and Madesis, P. (2007) DNA replication, recombination and repair in plastids. in *Cell and Molecular Biology of Plastids. Topics in Current Genetics, Vol. 19* (Bock, R. ed.), Springer, Berlin Heidelberg. pp 65-119

19. Elo, A., Lyznik, A., Gonzalez, D. O., Kachman, S. D., and Mackenzie, S. A. (2003) *Plant Cell* **15**, 1619-1631
20. Kubo, T., and Newton, K. J. (2008) *Mitochondrion* **8**, 5-14
21. Fauron, C., Allen, J., Clifton, S., and Newton, K. (2004) Plant mitochondrial genomes. in *Molecular Biology and Biotechnology of Plant Organelles* (Daniell, H., and Chase, C. eds.), Springer, Dordrecht. pp 151-177
22. Unseld, M., Marienfeld, J. R., Brandt, P., and Brennicke, A. (1997) *Nature Genetics* **15**, 57-61
23. Handa, H. (2003) *Nucleic Acids Res* **31**, 5907-5916
24. Notsu, Y., Masood, S., Nishikawa, T., Kubo, N., Akiduki, G., Nakazono, M., Hirai, A., and Kadowaki, K. (2002) *Mol Genet Genomics* **268**, 434-445
25. Kubo, T., Nishizawa, S., Sugawara, A., Itchoda, N., Estiati, A., and Mikami, T. (2000) *Nucleic Acids Res* **28**, 2571-2576
26. Clifton, S. W., Minx, P., Fauron, C. M., Gibson, M., Allen, J. O., Sun, H., Thompson, M., Barbazuk, W. B., Kanuganti, S., Tayloe, C., Meyer, L., Wilson, R. K., and Newton, K. J. (2004) *Plant Physiol* **136**, 3486-3503
27. Ogiwara, Y., Yamazaki, Y., Murai, K., Kanno, A., Terachi, T., Shiina, T., Miyashita, N., Nasuda, S., Nakamura, C., Mori, N., Takumi, S., Murata, M., Futo, S., and Tsunewaki, K. (2005) *Nucleic Acids Res* **33**, 6235-6250
28. Dai, H., Lo, Y. S., Litvinchuk, A., Wang, Y. T., Jane, W. N., Hsiao, L. J., and Chiang, K. S. (2005) *Nucleic Acids Res* **33**, 4725-4739
29. Sakai, A., Takano, H., and Kuroiwa, T. (2004) *Int Rev Cytol* **238**, 59-118
30. Lonsdale, D. M., Brears, T., Hodge, T. P., Melville, S. E., and Rottmann, W. H. (1988) *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **319**, 149-163
31. Bendich, A. J. (1996) *J Mol Biol* **255**, 564-588
32. Oldenburg, D. J., and Bendich, A. J. (1996) *Plant Cell* **8**, 447-461
33. Backert, S., Nielsen, B. L., and Börner, T. (1997) *Trends Plant Sci* **2**, 477-483
34. Marechal, A., and Brisson, N. (2010) *New Phytol* **186**, 299-317
35. Woloszyńska, M. (2010) *J Exp Bot* **61**, 657-671
36. Small, I. D., Isaac, P. G., and Leaver, C. J. (1987) *EMBO J* **6**, 865-869

37. Fauron, C., Casper, M., Gao, Y., and Moore, B. (1995) *Trends Genet* **11**, 228-235
38. Arrieta-Montiel, M., Lyznik, A., Woloszynska, M., Janska, H., Tohme, J., and Mackenzie, S. (2001) *Genetics* **158**, 851-864
39. André, C., Levy, A. A., and Walbot, V. (1992) *Trends Genet* **8**, 128-132
40. Marienfeld, J. R., Unseld, M., Brandt, P., and Brennicke, A. (1997) *Biol Chem* **378**, 859-862
41. Kmiec, B., Woloszynska, M., and Janska, H. (2006) *Curr Genet* **50**, 149-159
42. Allen, J. O., Fauron, C. M., Minx, P., Roark, L., Oddiraju, S., Lin, G. N., Meyer, L., Sun, H., Kim, K., Wang, C., Du, F., Xu, D., Gibson, M., Cifrese, J., Clifton, S. W., and Newton, K. J. (2007) *Genetics* **177**, 1173-1192
43. Wolfe, K. H., Li, W. H., and Sharp, P. M. (1987) *Proc Natl Acad Sci U S A* **84**, 9054-9058
44. Palmer, J. D., and Herbon, L. A. (1988) *J Mol Evol* **28**, 87-97
45. Abdelnoor, R. V., Yule, R., Elo, A., Christensen, A. C., Meyer-Gauen, G., and Mackenzie, S. A. (2003) *Proc Natl Acad Sci USA* **100**, 5968-5973
46. Zaegel, V., Guermann, B., Le Ret, M., Andrés, C., Meyer, D., Erhardt, M., Canaday, J., Gualberto, J. M., and Imbault, P. (2006) *Plant Cell* **18**, 3548-3563
47. Shedge, V., Arrieta-Montiel, M., Christensen, A. C., and Mackenzie, S. A. (2007) *Plant Cell* **19**, 1251-1264
48. Odahara, M., Kuroiwa, H., Kuroiwa, T., and Sekine, Y. (2009) *Plant Cell* **21**, 1182-1194
49. Arrieta-Montiel, M. P., Shedge, V., Davila, J., Christensen, A. C., and Mackenzie, S. A. (2009) *Genetics* **183**, 1261-1268
50. Rédei, G. P. (1973) *Mutat Res* **18**, 149-162
51. Sandhu, A. P., Abdelnoor, R. V., and Mackenzie, S. A. (2007) *Proc Natl Acad Sci USA* **104**, 1766-1770
52. Shedge, V., Davila, J., Arrieta-Montiel, M. P., Mohammed, S., and Mackenzie, S. A. (2010) *Plant Physiol* **152**, 1960-1970
53. Maier, R. M., and Schmitz-Linneweber, C. (2004) Plastid genomes. in *Molecular Biology and Biotechnology of Plant Organelles* (Daniell, H., and Chase, C. eds.), Springer, Dordrecht. pp 115-150

54. Brouard, J. S., Otis, C., Lemieux, C., and Turmel, M. (2010) *Genome Biol Evol* **2**, 240-256
55. Newton, K. J., Stern, D. B., and Gabay-Laughnan, S. (2009) Mitochondria and Chloroplasts. in *Handbook of Maize* (J.L., B., and S.C., H. eds.), Springer, Dordrecht. pp 481-504
56. Palmer, J. D. (1990) *Trends Genet* **6**, 115-120
57. Ogiwara, Y., Terachi, T., and Sasakuma, T. (1988) *Proc Natl Acad Sci USA* **85**, 8573-8577
58. Ogiwara, Y., and Ohsawa, T. (2002) *Genome* **45**, 956-962
59. Kolodner, R., and Tewari, K. K. (1972) *J Biol Chem* **247**, 6355-6364
60. Bendich, A. J. (2004) *Plant Cell* **16**, 1661-1666
61. Oldenburg, D. J., and Bendich, A. J. (2004) *J Mol Biol* **335**, 953-970
62. Sato, N., Terasawa, K., Miyajima, K., and Kabeya, Y. (2003) *Int Rev Cytol* **232**, 217-262
63. Kobayashi, T., Takahara, M., Miyagishima, S. Y., Kuroiwa, H., Sasaki, N., Ohta, N., Matsuzaki, M., and Kuroiwa, T. (2002) *Plant Cell* **14**, 1579-1589
64. Sato, N., Ohshima, K., Watanabe, A., Ohta, N., Nishiyama, Y., Joyard, J., and Douce, R. (1998) *Plant Cell* **10**, 859-872
65. Sato, N., Nakayama, M., and Hase, T. (2001) *FEBS Lett* **487**, 347-350
66. Chi-Ham, C. L., Keaton, M. A., Cannon, G. C., and Heinhorst, S. (2002) *Plant Mol Biol* **49**, 621-631
67. Sekine, K., Fujiwara, M., Nakayama, M., Takao, T., Hase, T., and Sato, N. (2007) *FEBS J* **274**, 2054-2069
68. Sekine, K., Hase, T., and Sato, N. (2002) *J Biol Chem* **277**, 24399-24404
69. Cannon, G. C., Ward, L. N., Case, C. I., and Heinhorst, S. (1999) *Plant Mol Biol* **39**, 835-845
70. Kang, Y. W., Lee, J. Y., Jeon, Y., Cheong, G. W., Kim, M., and Pai, H. S. (2010) *Plant Mol Biol* **72**, 569-583
71. Nakano, T., Murakami, S., Shoji, T., Yoshida, S., Yamada, Y., and Sato, F. (1997) *Plant Cell* **9**, 1673-1682
72. Murakami, S., Kondo, Y., Nakano, T., and Sato, F. (2000) *FEBS Lett* **468**, 15-18

73. Kato, Y., Murakami, S., Yamamoto, Y., Chatani, H., Kondo, Y., Nakano, T., Yokota, A., and Sato, F. (2004) *Planta* **220**, 97-104
74. Jeong, S. Y., Rose, A., and Meier, I. (2003) *Nucleic Acids Res* **31**, 5175-5185
75. Meier, I., Phelan, T., Gruissem, W., Spiker, S., and Schneider, D. (1996) *Plant Cell* **8**, 2105-2115
76. Samaniego, R., Jeong, S. Y., Meier, I., and de la Espina, S. M. (2006) *Planta* **223**, 1201-1206
77. Jeong, S. Y., Pfeffer, N., and Meier, I. (2004) *Planta* **219**, 298-302
78. Terasawa, K., Odahara, M., Kabeya, Y., Kikugawa, T., Sekine, Y., Fujiwara, M., and Sato, N. (2007) *Mol Biol Evol* **24**, 699-709
79. Sato, N., Albrieux, C., Joyard, J., Douce, R., and Kuroiwa, T. (1993) *EMBO J* **12**, 555-561
80. Pfalz, J., Liere, K., Kandlbinder, A., Dietz, K. J., and Oelmüller, R. (2006) *Plant Cell* **18**, 176-197
81. Hoeijmakers, J. H. (2001) *Nature* **411**, 366-374
82. Burhans, W. C., and Weinberger, M. (2007) *Nucleic Acids Res* **35**, 7545-7556
83. Rhoads, D. M., Umbach, A. L., Subbaiah, C. C., and Siedow, J. N. (2006) *Plant Physiol* **141**, 357-366
84. Apel, K., and Hirt, H. (2004) *Annu Rev Plant Biol* **55**, 373-399
85. Wyman, C., and Kanaar, R. (2006) *Annu Rev Genet* **40**, 363-383
86. Keeney, S. (2001) *Curr Top Dev Biol* **52**, 1-53
87. Friedberg, E. C. (2005) *Nat Rev Mol Cell Biol* **6**, 943-953
88. Friedberg, E. C., Fischhaber, P. L., and Kisker, C. (2001) *Cell* **107**, 9-12
89. Garcia-Diaz, M., and Bebenek, K. (2007) *Crit Rev Plant Sci* **26**, 105-122
90. Cline, S. D., and Hanawalt, P. C. (2003) *Nat Rev Mol Cell Biol* **4**, 361-372
91. Gutman, B. L., and Niyogi, K. K. (2009) *J Biol Chem* **284**, 17006-17012
92. Boesch, P., Ibrahim, N., Paulus, F., Cosset, A., Tarasenko, V., and Dietrich, A. (2009) *Nucleic Acids Res* **37**, 5690-5700
93. Nospikel, T. (2008) *DNA Repair (Amst)* **7**, 1155-1167
94. Jiricny, J. (2006) *Nat Rev Mol Cell Biol* **7**, 335-346
95. Abdelnoor, R. V., Christensen, A. C., Mohammed, S., Munoz-Castillo, B., Moriyama,

- H., and Mackenzie, S. A. (2006) *J Mol Evol* **63**, 165-173
96. Sonoda, E., Hocegger, H., Saberi, A., Taniguchi, Y., and Takeda, S. (2006) *DNA Repair (Amst)* **5**, 1021-1029
 97. Mahaney, B. L., Meek, K., and Lees-Miller, S. P. (2009) *Biochem J* **417**, 639-650
 98. Shuman, S., and Glickman, M. S. (2007) *Nat Rev Microbiol* **5**, 852-861
 99. Odom, O. W., Baek, K. H., Dani, R. N., and Herrin, D. L. (2008) *Plant J* **53**, 842-853
 100. Kohl, S., and Bock, R. (2009) *Plant J* **58**, 423-436
 101. Kass, E., and Jasin, M. (2010) *FEBS Lett* **584**, 3703-3708
 102. McVey, M., and Lee, S. E. (2008) *Trends Genet* **24**, 529-538
 103. Lee, K., and Lee, S. E. (2007) *Genetics* **176**, 2003-2014
 104. Kanno, A., Watanabe, N., Nakamura, I., and Hirai, A. (1993) Variations in chloroplast dna from rice (*Oryza-sativa*) - differences between deletions mediated by short direct-repeat sequences within a single-species. in *Theor Appl Genet*
 105. Moeykens, C. A., Mackenzie, S. A., and Shoemaker, R. C. (1995) *Plant Mol Biol* **29**, 245-254
 106. Kwon, T., Huq, E., and Herrin, D. L. (2010) *Proc Natl Acad Sci U S A* **107**, 13954-13959
 107. Lovett, S. T., Hurley, R. L., Suter, V. A., Jr., Aubuchon, R. H., and Lebedeva, M. A. (2002) *Genetics* **160**, 851-859
 108. Liskay, R. M., Letsou, A., and Stachelek, J. L. (1987) *Genetics* **115**, 161-167
 109. Cerutti, H., Osman, M., Grandoni, P., and Jagendorf, A. T. (1992) *Proc Natl Acad Sci USA* **89**, 8068-8072
 110. Inouye, T., Odahara, M., Fujita, T., Hasebe, M., and Sekine, Y. (2008) *Biosci Biotechnol Biochem* **72**, 1340-1347
 111. Khazi, F. R., Edmondson, A. C., and Nielsen, B. L. (2003) *Mol Genet Genomics* **269**, 454-463
 112. Manchekar, M., Scissum-Gunn, K., Song, D., Khazi, F., McLean, S. L., and Nielsen, B. L. (2006) *J Mol Biol* **356**, 288-299
 113. Symington, L. S. (2002) *Microbiol Mol Biol Rev* **66**, 630-670, table of contents
 114. Errico, A., and Costanzo, V. (2010) *EMBO Rep* **11**, 270-278
 115. Llorente, B., Smith, C. E., and Symington, L. S. (2008) *Cell Cycle* **7**, 859-864

116. Hastings, P. J., Ira, G., and Lupski, J. R. (2009) *PLoS Genet* **5**, e1000327
117. Hastings, P. J., Lupski, J. R., Rosenberg, S. M., and Ira, G. (2009) *Nat Rev Genet* **10**, 551-564
118. Zhang, F., Khajavi, M., Connolly, A. M., Towne, C. F., Batish, S. D., and Lupski, J. R. (2009) *Nat Genet*
119. Smith, C. E., Llorente, B., and Symington, L. S. (2007) *Nature* **447**, 102-105
120. Horvath, M. P. (2008) Single-stranded Nucleic Acid-binding Proteins. in *Protein-Nucleic Acid Interactions: Structural Biology* (Rice, P. A., and Correll, C. C. eds.), Royal Society of Chemistry, London. pp 91-128
121. Theobald, D. L., Mitton-Fry, R. M., and Wuttke, D. S. (2003) *Annual review of biophysics and biomolecular structure* **32**, 115-133
122. Bochkarev, A., and Bochkareva, E. (2004) *Curr Opin Struct Biol* **14**, 36-42
123. Arcus, V. (2002) *Curr Opin Struct Biol* **12**, 794-801
124. Ellenberger, T., and Tomkinson, A. E. (2008) *Annu Rev Biochem* **77**, 313-338
125. Maris, C., Dominguez, C., and Allain, F. H. (2005) *FEBS J* **272**, 2118-2131
126. Lunde, B. M., Moore, C., and Varani, G. (2007) *Nat Rev Mol Cell Biol* **8**, 479-490
127. Cléry, A., Blatter, M., and Allain, F. H. (2008) *Curr Opin Struct Biol* **18**, 290-298
128. Valverde, R., Edwards, L., and Regan, L. (2008) *FEBS J* **275**, 2712-2726
129. Siomi, H., Matunis, M. J., Michael, W. M., and Dreyfuss, G. (1993) *Nucleic Acids Res* **21**, 1193-1198
130. Grishin, N. V. (2001) *Nucleic Acids Res* **29**, 638-643
131. Wilusz, C. J., and Wilusz, J. (2005) *Nat Struct Mol Biol* **12**, 1031-1036
132. Thomas, M. C., and Chiang, C.-M. (2006) *Crit Rev Biochem Mol Biol* **41**, 105-178
133. Brandsen, J., Werten, S., van der Vliet, P. C., Meisterernst, M., Kroon, J., and Gros, P. (1997) *Nat Struct Biol* **4**, 900-903
134. Werten, S., and Moras, D. (2006) *Nat Struct Mol Biol* **13**, 181-182
135. Jonker, H. R. A., Wechselberger, R. W., Boelens, R., Folkers, G. E., and Kaptein, R. (2005) *Biochemistry* **44**, 827-839
136. Batta, K., and Kundu, T. K. (2007) *Mol Cell Biol* **27**, 7603-7614
137. Rajagopalan, S., Andreeva, A., Teufel, D. P., Freund, S. M., and Fersht, A. R. (2009) *J Biol Chem* **284**, 21728-21737

- 138. Jonker, H. R. A., Wechselberger, R. W., Boelens, R., Kaptein, R., and Folkers, G. E. (2006) *Biochemistry* **45**, 5067-5081
- 139. Jonker, H. R. A., Wechselberger, R. W., Pinkse, M., Kaptein, R., and Folkers, G. E. (2006) *FEBS J* **273**, 1430-1444
- 140. Desveaux, D., Allard, J., Brisson, N., and Sygusch, J. (2002) *Nat Struct Biol* **9**, 512-517
- 141. Schumacher, M. A., Karamooz, E., Zíková, A., Trantírek, L., and Lukes, J. (2006) *Cell* **126**, 701-711
- 142. Graebisch, A., Roche, S., and Niessing, D. (2009) *Proc Natl Acad Sci USA*
- 143. Xu, Q., Traag, B., Willemse, J., McMullan, D., Miller, M., Elsliger, M., Abdubek, P., Astakhova, T., Axelrod, H., Bakolitsa, C., Carlton, D., Chen, C., Chiu, H., Chruszcz, M., Clayton, T., Das, D., Deller, M., Duan, L., Ellrott, K., Ernst, D., Farr, C., Feuerhelm, J., Grant, J., Grzechnik, A., Grzechnik, S., Han, G., Jaroszewski, L., Jin, K., Klock, H., Knuth, M., Kozbial, P., Krishna, S., Kumar, A., Marciano, D., Minor, W., Mommaas, A., Morse, A., Nigoghossian, E., Nopakun, A., Okach, L., Oommachen, S., Paulsen, J., Puckett, C., Reyes, R., Rife, C., Sefcovic, N., Tien, H., Trame, C., van den Bedem, H., Wang, S., Weekes, D., Hodgson, K., Wooley, J., Deacon, A., Godzik, A., Lesley, S., Wilson, I., and van Wezel, G. (2009) *J Biol Chem* **284**, 25268-25279
- 144. Gallia, G. L., Johnson, E. M., and Khalili, K. (2000) *Nucleic Acids Res* **28**, 3197-3205
- 145. Auweter, S. D., Oberstrass, F. C., and Allain, F. H. (2006) *Nucleic Acids Res* **34**, 4943-4959
- 146. Coulocheri, S. A., Pigis, D. G., Papavassiliou, K. A., and Papavassiliou, A. G. (2007) *Biochimie* **89**, 1291-1303
- 147. Luscombe, N. M., Laskowski, R. A., and Thornton, J. M. (2001) *Nucleic Acids Res* **29**, 2860-2874
- 148. Theobald, D. L., and Schultz, S. C. (2003) *EMBO J* **22**, 4314-4324
- 149. Lei, M., Podell, E. R., and Cech, T. R. (2004) *Nat Struct Mol Biol* **11**, 1223-1229
- 150. Bochkarev, A., Pfuetzner, R. A., Edwards, A. M., and Frappier, L. (1997) *Nature* **385**, 176-181
- 151. Ferrari, M. E., Bujalowski, W., and Lohman, T. M. (1994) *J Mol Biol* **236**, 106-123
- 152. Raghunathan, S., Kozlov, A. G., Lohman, T. M., and Waksman, G. (2000) *Nat Struct*

Biol **7**, 648-652

153. Buczek, P., and Horvath, M. P. (2006) *J Biol Chem* **281**, 40124-40134
154. Buczek, P., and Horvath, M. P. (2006) *J Mol Biol* **359**, 1217-1234
155. Chen, Z., Yang, H., and Pavletich, N. P. (2008) *Nature* **453**, 489-484
156. Meyer, R. R., and Laine, P. S. (1990) *Microbiol Rev* **54**, 342-380
157. Shereda, R. D., Reiter, N. J., Butcher, S. E., and Keck, J. L. (2009) *J Mol Biol* **386**, 612-625
158. Lu, D., and Keck, J. L. (2008) *Proc Natl Acad Sci USA* **105**, 9169-9174
159. Roy, R., Kozlov, A., Lohman, T., and Ha, T. (2009) *Nature*
160. Edmondson, A. C., Song, D., Alvarez, L. A., Wall, M. K., Almond, D., McClellan, D. A., Maxwell, A., and Nielsen, B. L. (2005) *Mol Genet Genomics* **273**, 115-122
161. Matton, D. P., Prescott, G., Bertrand, C., Camirand, A., and Brisson, N. (1993) *Plant Mol Biol* **22**, 279-291
162. Desveaux, D., Després, C., Joyeux, A., Subramaniam, R., and Brisson, N. (2000) *Plant Cell* **12**, 1477-1489
163. Boyle, B., and Brisson, N. (2001) *Plant Cell* **13**, 2525-2537
164. Desveaux, D., Subramaniam, R., Després, C., Mess, J. N., Lévesque, C., Fobert, P. R., Dangel, J. L., and Brisson, N. (2004) *Dev Cell* **6**, 229-240
165. Ni, C. Z., Wang, H. Q., Xu, T., Qu, Z., and Liu, G. Q. (2005) *Cell Res* **15**, 725-733
166. XuYan, L., HaQing, W., Tao, X., QinHong, C., DongTao, R., and GuoQin, L. (2007) *Chinese Sci Bull* **52**, 1338-1346
167. Lai, C., Xiong, J., Li, X., and Qin, X. (2009) *Plant Cell Rep* **28**, 851-860
168. Xiong, J. Y., Lai, C. X., Qu, Z., Yang, X. Y., Qin, X. H., and Liu, G. Q. (2009) *Plant Mol Biol* **71**, 437-449
169. Kwon, C., and Chung, I. K. (2004) *J Biol Chem* **279**, 12812-12818
170. Yoo, H. H., Kwon, C., Lee, M. M., and Chung, I. K. (2007) *Plant J* **49**, 442-451
171. Prikryl, J., Watkins, K. P., Friso, G., Wijk, K. J., and Barkan, A. (2008) *Nucleic Acids Res* **36**, 5152-5165
172. Melonek, J., Mulisch, M., Schmitz-Linneweber, C., Grabowski, E., Hensel, G., and Krupinska, K. (2010) *Planta* **232**, 471-481
173. Maréchal, A., Parent, J. S., Véronneau-Lafortune, F., Joyeux, A., Lang, B. F., and

- Brisson, N. (2009) *Proc Natl Acad Sci USA* **106**, 14693-14698
174. Krause, K., Kilbiński, I., Mulisch, M., Rödiger, A., Schäfer, A., and Krupinska, K. (2005) *FEBS Lett* **579**, 3707-3712
175. Chalker, D. L. (2008) *Biochim Biophys Acta* **1783**, 2130-2136
176. Saha, S., and Kapler, G. M. (2000) *J Mol Biol* **295**, 423-439
177. Saha, S., Nicholson, A., and Kapler, G. M. (2001) *J Biol Chem* **276**, 45417-45426
178. Mohammad, M., York, R. D., Hommel, J., and Kapler, G. M. (2003) *Mol Cell Biol* **23**, 5005-5017
179. Yakisich, J. S., Sandoval, P. Y., Morrison, T. L., and Kapler, G. M. (2006) *Mol Biol Cell* **17**, 5185-5197
180. Desveaux, D., Allard, J., Brisson, N., and Sygusch, J. (2002) *Acta Crystallogr D Biol Crystallogr* **58**, 296-298
181. Desveaux, D., Maréchal, A., and Brisson, N. (2005) *Trends Plant Sci* **10**, 95-102
182. Schoeffler, A. J., and Berger, J. M. (2005) *Biochem Soc Trans* **33**, 1465-1470
183. Sugino, A., Peebles, C. L., Kreuzer, K. N., and Cozzarelli, N. R. (1977) *Proc Natl Acad Sci U S A* **74**, 4767-4771
184. Chen, C. R., Malik, M., Snyder, M., and Drlica, K. (1996) *J Mol Biol* **258**, 627-637
185. Snyder, M., and Drlica, K. (1979) *J Mol Biol* **131**, 287-302
186. Gellert, M., Mizuuchi, K., O'Dea, M. H., Itoh, T., and Tomizawa, J. I. (1977) *Proc Natl Acad Sci U S A* **74**, 4772-4776
187. Wall, M. K., Mitchenall, L. A., and Maxwell, A. (2004) *Proc Natl Acad Sci USA* **101**, 7821-7826
188. Rupp, B. (2010) *Biomolecular Crystallography*, Garland Science, New York
189. Page, R. (2008) *Methods Mol Biol* **426**, 345-362
190. Goldman, N., Thorne, J. L., and Jones, D. T. (1998) *Genetics* **149**, 445-458
191. Hui, R., and Edwards, A. (2003) *J Struct Biol* **142**, 154-161
192. Vermel, M., Guermann, B., Delage, L., Grienemberger, J. M., Maréchal-Drouard, L., and Gualberto, J. M. (2002) *Proc Natl Acad Sci USA* **99**, 5866-5871
193. Maréchal, A., Parent, J. S., Sabar, M., Véronneau-Lafortune, F., Abou-Rached, C., and Brisson, N. (2008) *BMC Plant Biol* **8**, 42
194. Krissinel, E., and Henrick, K. (2007) *J Mol Biol* **372**, 774-797

195. Zhang, Y., Raudah, S., Teo, H., Teo, G. W., Fan, R., Sun, X., and Orner, B. P. (2010) *J Biol Chem* **285**, 12078-12086
196. Mapelli, M., Muhleisen, M., Persico, G., van Der Zandt, H., and Tucker, P. A. (2000) *J Virol* **74**, 8812-8822
197. Holm, L., and Park, J. (2000) *Bioinformatics* **16**, 566-567
198. Sugiman-Marangos, S., and Junop, M. S. (2010) *Nucleic Acids Res* **38**, 3432-3440
199. Srivastava, S., and Moraes, C. T. (2001) *Hum Mol Genet* **10**, 3093-3099
200. Morel, F., Renoux, M., Lachaume, P., and Alziari, S. (2008) *Mutat Res* **637**, 111-117
201. Rowan, B. A., Oldenburg, D. J., and Bendich, A. J. (2010) *J Exp Bot* **61**, 2575-2588
202. Shereda, R. D., Kozlov, A. G., Lohman, T. M., Cox, M. M., and Keck, J. L. (2008) *Crit Rev Biochem Mol Biol* **43**, 289-318
203. Backert, S., Lurz, R., Oyarzabal, O. A., and Börner, T. (1997) *Plant Mol Biol* **33**, 1037-1050
204. Chase, J. W., and Williams, K. R. (1986) *Annu Rev Biochem* **55**, 103-136
205. Falkenberg, M., Larsson, N. G., and Gustafsson, C. M. (2007) *Annu Rev Biochem* **76**, 679-699
206. Oliveira, M. T., and Kaguni, L. S. (2010) *PLoS One* **5**, e15379
207. Farr, C. L., Matsushima, Y., Lagina, A. T., 3rd, Luo, N., and Kaguni, L. S. (2004) *J Biol Chem* **279**, 17047-17053
208. Broderick, S., Rehmet, K., Concannon, C., and Nasheuer, H. P. (2010) *Subcell Biochem* **50**, 143-163
209. Grove, A. (2010) *Curr Issues Mol Biol* **13**, 1-12
210. Kamashev, D., Balandina, A., Mazur, A. K., Arimondo, P. B., and Rouviere-Yaniv, J. (2008) *Nucleic Acids Res* **36**, 1026-1036
211. Janner, A. (2008) *Acta Crystallogr, A, Found Crystallogr* **64**, 494-502
212. Krojer, T., Sawa, J., Schäfer, E., Saibil, H. R., Ehrmann, M., and Clausen, T. (2008) *Nature* **453**, 885-890
213. Kim, K. K., Kim, R., and Kim, S. H. (1998) *Nature* **394**, 595-599
214. Urich, T., Gomes, C. M., Kletzin, A., and Frazão, C. (2006) *Science* **311**, 996-1000
215. Glynn, S. E., Baker, P. J., Sedelnikova, S. E., Davies, C. L., Eadsforth, T. C., Levy, C. W., Rodgers, H. F., Blackburn, G. M., Hawkes, T. R., Viner, R., and Rice, D. W.

- (2005) *Structure* **13**, 1809-1817
216. Theil, E. C. (1987) *Annu Rev Biochem* **56**, 289-315
217. Grant, R. A., Filman, D. J., Finkel, S. E., Kolter, R., and Hogle, J. M. (1998) *Nat Struct Biol* **5**, 294-303
218. Pfannschmidt, T. (2010) *Trends Plant Sci* **15**, 427-435